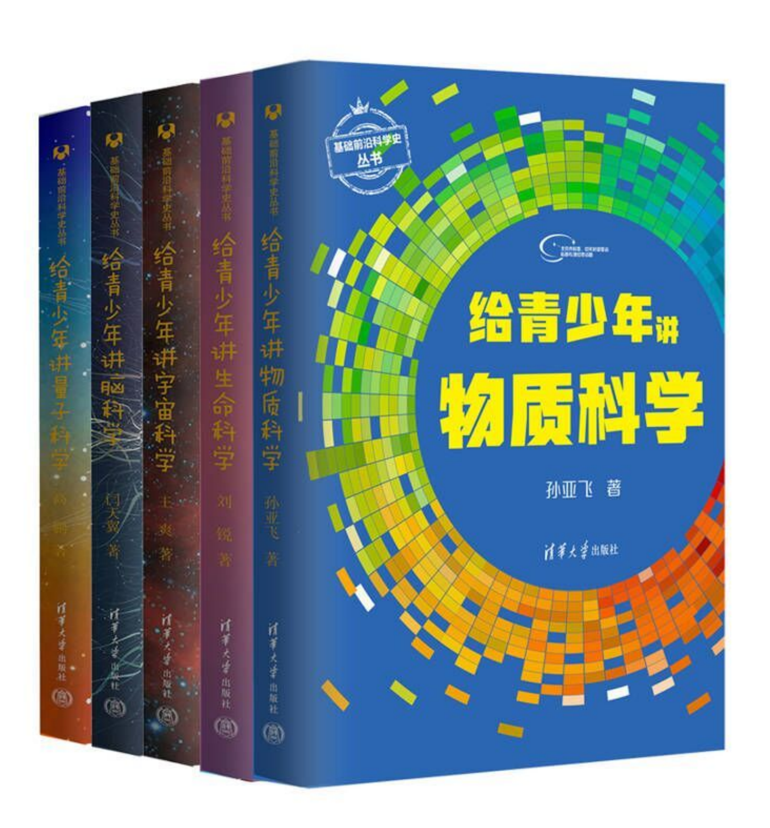


基础前沿科学史丛书
给青少年讲物质科学
孙亚飞 著
清华大学出版社
给青少年讲生命科学
刘锐 著
给青少年讲宇宙科学
王爽 著
给青少年讲脑科学
给青少年讲量子科学

# 总目录



基础前沿科学史
丛书
北京市科委、中关村管委会
科普专项经费资助
给青少年讲
物质科学
孙亚飞 著
清华大学出版社

基础前沿科学史丛书

# 给青少年讲物质科学

孙亚飞　著

清华大学出版社
北　京

图书在版编目（**CIP**）数据

给青少年讲物质科学 / 孙亚飞著 .—北京：清华大学出版社，2022.11

（基础前沿科学史丛书）

ISBN 978-7-302-62140-9

Ⅰ . ①给…　Ⅱ . ①孙…　Ⅲ . ①物质 - 青少年读物　Ⅳ . ①O4-49

中国版本图书馆CIP数据核字（2022）第204635号

责任编辑：刘　杨

封面设计：意匠文化·丁奔亮

责任校对：王淑云

责任印制：沈　露

出版发行：清华大学出版社

网　　址：http://www.tup.com.cn, http://www.wqbook.com

地　　址：北京清华大学学研大厦A座

邮　　编：100084

社总机：010-83470000

邮　　购：010-62786544

投稿与读者服务：010-62776969，c-service@tup.tsinghua.edu.cn

质量反馈：010-62772015，zhiliang@tup.tsinghua.edu.cn

印装者：三河市龙大印装有限公司
经　　销：全国新华书店
开　　本：165mm×235mm
印　　张：11
字　　数：119千字
版　　次：2022年12月第1版
印　　次：2022年12月第1次印刷
定　　价：55.00元

产品编号：097618-01

# 丛书序　给面向青少年的科普出版点一把新火

2022年是《中华人民共和国科普法》通过的第20年，在这样一个对科普工作意义不凡的年份，由北京市科学技术委员会（以下简称市科委）发起，清华大学出版社组织的“基础前沿科学史丛书”正式出版了。这套书给面向青少年的科普出版点了一把新火。

2022年9月4日，中共中央办公厅、国务院办公厅印发《关于新时代进一步加强科学技术普及工作的意见》，进一步强调“科学技术普及是国家和社会普及科学技术知识、弘扬科学精神、传播科学思想、倡导科学方法的活动，是实现创新发展的重要基础性工作”。科学技术普及是科技知识、科学精神、科学思想、科学方法的薪火相传——是“薪火”，也是“新火”。

市科委搭台，出版社唱戏，这套书给面向青少年的科普图书出版模式点了一把新火。市科委于2021年11月发布了“创作出版‘基础前沿科学史’系列精品科普图书”的招标公告，明确要求中标方在一年的时间内，以物质科学、生命科学、宇宙科学、脑科学、量子科学为主题，组织“基础前沿科学史”系列精品科普图书（共5册）出版工作；同步设计制作科普电子书；通过网络媒体对图书进行宣传推广等服务内容。这些服务内容以融合出版为基础，以社会效益为初心。服务内容的短短几句话，每一句背后都是特别繁复的工作内容。想在一年的时间内，尤其是在2022年新冠肺炎疫情期间，完成这些工作的难度可想而知，然而秉承“自强不息，厚德载物”的清华大学出版社的出版团队做到了。

中国科学家，讲好中国故事，这套书给面向青少年的科普图书选题内容点了一把新火。中国特色社会主义进入新时代，新一轮科技革命和产业变革正在深入发展，基础前沿科学改变着人们的生产生活方式及思维模式。《中华人民共和国国民经济和社会发展第十四个五年规划和2035年远景目标纲要》提出：在事关国家安全和发展全局的基

础核心领域，制定实施战略性科学计划和科学工程。物质科学、生命科学、宇宙科学、脑科学、量子科学等领域，迫切需要更多人才参与研究，而前沿科学人才的建设培养，要从青少年抓起。这5本书的作者都是中国本土从事相关专业领域工作的科学家，这5本书都是他们依托自己工作进行的原创性工作。虽然内容必然涉及科学史的内容，但中国科学家尤其是近些年的贡献也得到了充分展示。

初心教育，润物无声，这套书给面向青少年的科普图书科普创作点了一把新火。习近平总书记提出：科技创新、科学普及是实现创新发展的两翼，要把科学普及放在与科技创新同等重要的位置。因此，针对前沿科技领域知识的科普成为重点。如何创作广受青少年欢迎的优秀科普图书，充分发挥科普图书的媒介作用，帮助青少年树立投身前沿科学领域的梦想，是当前科普出版工作的重点之一，这对具体的科普创作方法提出了要求。这套书，看得出来在创作之初即统一了整体创作思路，在作者进行具体创作时又保持了自己的语言习惯和科普风格。这套书充分体现了，面向青少年的科普图书创作，应该循序渐进，张弛有度，绘声绘色，娓娓道来，以科学家的故事吸引他们，温故科学家的研究之路，知新科学家的科研理念，以科学精神润物细无声。

靡不有初，鲜克有终。2022年10月16日，习近平总书记在中国共产党第二十次全国代表大会报告中强调“教育、科技、人才是全面建设社会主义现代化国家的基础性、战略性支撑”。且将新火试新茶，诗酒趁年华。希望清华大学出版社的这套“基础前沿科学史丛书”为广大青少年推开科学技术事业的一扇门，帮助他们系好投身科学技术事业的第一粒扣子，在全面建设社会主义现代化强国的新征程上行稳致远。

中国工程院院士

清华大学教授



# 前言　给青少年的物质科学

小的时候，我生活在农村。那时的农村没有煤气灶，也舍不得用太多电，做饭最常用的锅就是土灶台，烧柴火给铁锅供热。灶台操作起来可是麻烦得很，光是生火就能生出各种意外。所以，家里大人做饭免不了还需要我打下手，别的忙我也帮不上，拉一拉风箱还是可以的。

曾记得，坐在灶边望着炉火，我拉动风箱，火苗就会猛地蹿一蹿，一不小心还会冲出灶膛，燎到额前的头发。每当这时，我都会感到脸上一热，但是习惯之后，倒也不觉得有什么危险，只是有些好奇。

我好奇的是，吹蜡烛的时候，一点小风就能把烛火吹灭，可是风箱施展出那么大的风以后，为什么火焰却会越吹越大？

外婆跟我说，这是因为屋里供了灶王爷——他是一位管灶火的神仙。有了灶王爷的保佑，炉灶里的三昧真火才会旺盛，而且还不会因此失火。

后来，看到《西游记》中孙悟空借来假芭蕉扇挥舞时，火焰山上的火越烧越大，我对三昧真火的这个说法深信不疑。可是，看到《三国演义》里赤壁之战东风助阵的场面，我又有些将信将疑了。

多年以后，我读到了拉瓦锡反驳“燃素”论的故事，对于这个问题的好奇总算告一段落。燃烧是一种氧化现象，通常需要氧气的参与，因此，当风箱给炉灶送风时，加大了空气压力，更多的氧气被一同送入了灶膛。于是，有了氧气的加持，火焰也会更旺盛。不过，风在刮来新鲜氧气的同时，却也带走了热量，若是火焰的温度因此而降低到燃点以下，火就可能会被吹灭。由此也就很容易明白，为什么风能吹灭小火，却会扬起大火。

对于童年时期的不解，用这样物质的思维去解释，或许有些许的无趣，不像“三昧真火”那么引人入胜，但它却是好奇的最佳伙伴。它不会让童心泯灭，只会带着这种“不安分”走向新的空间。

所以，青少年知道多一点物质科学，应该是件好事。

《给青少年讲物质科学》，是我第一次系统性地创作“大科学”的作品。这倒不是我妄自托大，实属物质科学自身的属性。任何一门自然科学都是研究物质的科学，只是研究的角度不同：物理学研究物质之间的作用力与能量；化学研究物质的组成、结构、反应与功能；生物学研究的是物质如何能够“活”起来；天文学研究物质诞生的“老家”……很多时候，关于物质的种种，并不只是单纯的科学命题，更是人类反复思索的哲学命题。

将这些命题完整地解答出来，是古往今来所有人共同的使命。哪怕已经觍着脸站到了巨人的肩膀上，我也只能记录其中亿万分之一二。

在物质世界形成的时空中，翻开这本由物质构成的小书，如果能够因此引发我们对物质的思考，那么物质世界也一定会因为我们的思考而改变。

# 目 录

# 1　物质是什么

2015年12月17日，中国酒泉卫星发射中心利用长征二号D型运载火箭成功地将“悟空”送上云霄，进入太空。“悟空”没有打上南天门，更没有去蟠桃盛宴赴会，只是在用它的“火眼金睛”——空间望远镜等各种精密仪器，以前所未有的灵敏度和能量范围搜索一些神秘的信号。“悟空”是一颗暗物质粒子探测卫星（Dark Matter Particle Exploer, DAMPE），也是中国科学院空间科学战略性先导科技专项中的第一颗空间天文卫星，如果能够如愿，那么“悟空”就能找到暗物质存在的更多证据，甚至有可能让我们一睹暗物质的“真容”。目前，它已在暗物质间接探测和宇宙线起源等方面作出了重要贡献，这标志着我国在空间高能粒子探测领域已跻身世界最前列。

在整个宇宙之中，我们可以看到的物质，大约只占所有物质的5%，剩下的那部分，大约27%是由暗物质构成，还有68%则是更让人摸不着头脑的暗能量。

早在1933年的时候，一位名叫兹威基·弗里茨（Zwicky Fritz，1898年—1974年）的天文学家在研究星系团内星系运动的过程中取得一系列的研究成果，预测了暗物质存在的可能性。当时，他用望远镜观察遥远的星系，却发现他所看到的星系，比通过计算得出的星系质量小得多。于是他就推断，在宇宙中一定还有很多我们看不见的物质，但我们可以感受到它们。就好像夏夜躲在地下鸣叫的蝼蛄那样，只有打开手电筒，挖出土坑，我们才有可能找到它们的踪迹。既然是躲在宇宙中暗处又不发光的物质，那就叫它们暗物质好了。

几十年过去了，人们一直在寻找这些躲在暗处的“蝼蛄”。

在技术水平不断提升的背景下，这似乎不是什么难题。如果要在黑夜中寻找什么，如今的我们并不总是需要等到白天来临，也不必开灯，用红外线夜视仪一样可以奏效。这是因为，能被人眼看到的光，只是被称为“可见光”的那部分，只有当物品发射或反射出可见光时，人眼才能看到它们。当黑夜来临之时，由于这些物体只能发出很微弱

的可见光，我们自然也就很难再看到它们。可是，即便是在暗处的物品，仍然可以发射出红外光，人的肉眼虽然看不见，可是戴上一双可以看到红外光的“眼睛”，就能看清黑暗中的世界了。我们探索未知世界需要动用很多技术，红外线只不过是一个缩影。我们想要找到的目标，哪怕就是像蝼蛄那样躲在地下，我们现在也有很多办法找到它们的踪迹。实际上，在自然科学中，常常把紫外线、可见光和红外线，统称为光辐射，它们都属于光波。



伽利略

正因为如此，人们一开始并没有把弗里茨预言的暗物质当回事，只是猜测，这些看不见的物质，大概就是一些光线暗淡的星球罢了，我们的肉眼看不到，是因为它们实在太遥远，说不定换个合适的仪器就能看到了。更何况，这样的故事早就已经发生过——人类在地球上原本看不到木星的那些卫星，可是伽利略用一台非常朴素的望远镜，就发现了其中的四颗，除了木卫二，其他三颗甚至比月球还要大。所以，关于暗物质的一个合理假设，是宇宙中必然有很多没有被我们发现的天体和星系等，它们发出的可见光太弱，而地球距离它们又实在太远，只有靠一些间接的办法才能找到它们。

后来，在仪器的帮助下，科学家们借助红外线及其他各种技术，果然找到了一些黑暗中的星球，验证了这个想法。事实上，只要是人类可以想到的办法，全都用上了，这才有了一些新的发现。但是，就算加上这些星球，还是有很多物质，我们依然把它们视为“暗物质”——可以感受到它们的存在，却没有任何办法观测。

会不会还有其他一些可能呢？随着研究的深入，科学家们否定了一个又一个假说，至于暗物质究竟是什么，到现在还是不知道。唯一达成共识的是，科学家认为，这些暗物质虽然包括不发光的天体，星系晕物质等重子暗物质，但我们一般意义上更关注的是那些仅参与引力作用、弱相互作用而不参与电磁作用的非重子中性粒子等。无论如何，我们相信，终有一天我们可以认识它，并由此拓展我们的视野。

如今，当我们问起“物质是什么”的时候，也只能就已知这5%的宇宙做出回答，就是那些具有客观实在性的物质。对于那些未知的暗物质，还有更神秘的暗能量，我们不敢妄言。

而在这些可以被观测的宇宙中，我们将一切都视为物质——除了我们对物质的理解本身，这种理解被称为“意识”。物质和意识之间的关系，是一个古老的哲学命题，它就像一个绕不过去的海角，当我们对这个世界有所思考的时候，总不免要在这处海角——被称为唯物主义哲学基石的物质概念——逗留，有时候还要写个“在此一游”，广而告之。围绕着物质与意识，人们分为若干门派争论不休，怎样看待它，也就决定了“物质是什么”的答案。

意识就好比是已知物质世界的边缘。如果我们把意识看成一个气球，那么这个气球就将物质世界分为两个部分，气球以外的那部分，我们不知道是什么，潜意识里感觉存在的那部分就称它们是暗物质，而在气球内的这部分，显意识感受到的是这5%已知的宇宙世界。只不过，每个人的意识各不相同，气球的大小也不尽相同，看待物质的角度也就有了巨大的差异。意识对物质的关心问题也是哲学的基本问题。

不经意间，对于物质的观点甚至会左右我们对宇宙的探索。

当美国科学家本杰明·富兰克林（Benjamin Franklin，1706年—1790年）在那个雷雨天放出风筝时，他把难以捉摸的闪电当成了物质。

当德国物理学家威廉·康拉德·伦琴（Wilhelm Conrad Rontgen，1845年—1923年）给太太拍摄手掌骨骼的照片时，他把神秘未知的X射线（伦琴射线，俗称X光）当成了物质。

当英国物理学家彼得·希格斯（Peter Higgs，1929年—）建立起希格斯场的假说时，他把未被观察到的希格斯玻色子当成了物质。

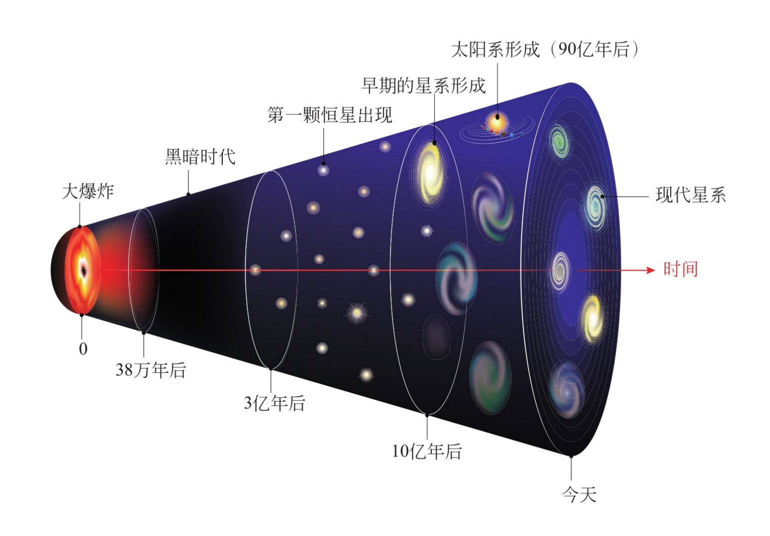
……

类似这样的故事或许永远都不会结束。只要我们合理而勇敢地去放大意识，似乎总有机会去找到一点新的物质。就像当我们把气球越吹越大时，那就意味着，气球外的空间又小了一点点。无论这点变化多么微不足道，它都意味着我们还在继续前行。

所以，物质就是不依赖我们的意识又能够被我们所理解的一切真实存在的事物，除了我们的“理解”本身。物质的集合一直在变化，哪怕我们不能就“物质是什么”的问题达成共识，也不必为此感到烦恼，这就是物质世界本来的模样。

至于我们熟悉的这个物质世界，要从138亿年前说起。那时，所有的物质，也包括所有的能量，全都集中在一个“点”上，这个点被称为奇点。

突然，奇点发生了爆炸，也就是著名的“宇宙大爆炸”。很快，物质以粒子的形式溅射出来，宇宙开始膨胀。此时，宇宙的温度高得出奇，电子、夸克还有胶子都是宇宙中稳定存在的粒子。它们都小得出奇，然而希格斯玻色子为它们赋予了质量。



宇宙演化示意图

膨胀的宇宙快速降温，夸克和胶子也开始相互碰撞，胶子在夸克之间传递着强相互作用，就像胶水一样把夸克粘接在一起，形成了更大的粒子。我们现在知道，夸克是组成强子的更基本的粒子，有6种夸克及对应的反夸克，其中有两种分别被称为上夸克（u）和下夸克（d）。它们都带有电荷：上夸克带有2/3个正电荷，下夸克则带有1/3个负电荷。于是，两个上夸克和一个下夸克结合成带有一个正电荷的粒子，它被称为质子；两个下夸克和一个上夸克则结合成电荷为零的粒子，它被称为中子。通过加速器实验，科学家已全部观测到这些夸克粒子的存在，但实验上还没能分离出单独存在的具有像这样分数电荷的夸克。

随着宇宙温度进一步降低，质子和中子也可以紧紧地结合在一起，如果它们再碰到带有一个负电荷的电子，就能够组成原子。原子是构成万物的基石，我们这个有关物质的故事也将从这里正式开启。

# 2　世界万物的基石——原子的概念是怎样被提出来的

## 被分割的物质

当人们对宇宙中的物质的认识发展到原子层面的时候，似乎一切故事都将变得简单而清晰，毕竟原子的世界，是一个我们很容易触摸到的物质世界。尽管我们早已明白，原子并非是最小的物质单位，但还是会将它看作构成这个世界的基础，因为它是组成单质和化合物分子的最小微粒。

然而，想象出“原子”这种模型，对人类来说无异于一场巨大的思想变革。

地壳表面的一块岩石，无论它有多结实，在水、生物与风力等因素的长期联合作用下也会发生崩解——这就是地质学所说的风化作用。这样的风化作用，它可以是物理的、也可以是化学的或生物的作用。例如，大石头逐渐风化成小石头，而小石头还可能会继续裂开，再被苔藓附着，伴随着生物的作用，变成砂砾，最后实在太过微小，和黏土揉在一起，不分彼此。

这样一种司空见惯的现象，会给人带来自然而然的启发：大块头的物质是由小个子物质组合而成，而小个子物质又是由更小的物质组成。

毫无意外，顺着这个思路，我们会进一步展开联想，向自然界发问：如果把石头这样的物质一直切割下去，是否存在最小的石头单元？把这种最小的石头单元堆砌起来，是否又可以重新变成石头？

很难说这些问题的实际价值如何，它们看起来像是我们现代人吃饱喝足才会拥有的意趣。至于使用石器的史前人类，在面对满地形状各异的石头时，他们是否也有过这样的想法，如今早已不得而知。

但是，随着人类文明逐渐建立，这些问题在生活中其实是难以回避的。

比方说，河东和河西有两个部落做买卖。河西边的部落有黄金，不妨就叫黄金部落；河东边的部落有贝壳，那就叫贝壳部落。黄金和贝壳这两种物品，都曾经被作为货币使用，所以，这两个部落都有了采购物品的资本。

有一天，黄金部落有个人带了一锭黄金去河东做生意，而这锭金子可以买一头大牛或者两头小牛犊。转悠了半天以后，他在贝壳部落看中了一头小牛犊，就想买下它。这牛犊只要半锭金子，于是黄金部落的这位买家就和牛主人商量，把金锭切一半，刚好可以买下小牛犊。对牛主人来说，这个办法似乎没有理由拒绝，他就顺理成章答应了。

过了几天，贝壳部落的这位牛主人也去黄金部落赶集了，他带了一枚稀有的贝壳，想去买个木犁回来耕地。巧得很，上次买牛犊的那个人是位木匠，刚好打了一把很不错的木犁，被牛主人瞧上了。但是，一枚贝壳能买下两把这样的木犁，于是牛主人就琢磨，要不和上次一样，把贝壳也一分为二，问题不就解决了？显然，对于黄金部落的木匠来说，他不太可能会答应这个要求，因为常识告诉他，被分为两半的贝壳不值钱。

到底是什么决定了黄金和贝壳各自的价值？从这两笔买卖中很容易看得出来，虽然黄金和贝壳都是货币，但是黄金的价值在于这种物质本身，和它的外形无关；而贝壳的价值却体现在物品之上，和它的外形有关。换句话说，把一锭黄金一分为二，得到的是两锭小一些的黄金；把贝壳一分为二，得到的却是贝壳的碎片，而非两个小一点的“黄金”。




可以分割的黄金

生活中还有比这更复杂的情况。

按照国家的相关规定，在图书的版权页上，出版机构会注明书的开本信息，比如本书在版权页开本那个地方写的就是“165mm×235mm”。我们稍加留心就会发现，不同开本的书，大小差异很大。一般来说，出版机构会根据书的内容来确定选用什么样的开本。除了图书，生活中我们还时常会见到报纸、便笺、作业本、日记本、海报等各式各样、大小不一的纸。这说明，一张纸，按需裁剪之后，便可以发挥相应的价值。相应地，如果不按需裁剪，就会变成废纸。

可见，纸张和黄金还有贝壳又不一样。它既不像黄金那样，无论怎么切割都能保留货币属性，但也不像贝壳那样，哪怕只是切成两部分都会一文不值。

黄金部落的那个木匠，总是要和各种木头打交道，而他在收集木头的时候，也会面临和纸张一样的问题，大一点的木头能做房梁，小一点的木头可以打木犁或作为各种家具部件，可要是把木头锯得太细碎，最后就只能当柴火烧了。

贝壳部落的那个耕农，要是牛犊子不幸夭折，等他含泪卖牛肉的时候，还会碰到更蹊跷的事情——同一块牛肉，分割成两块同样大小的肉之后，要是一块肥、一块瘦，就算大小相同，实际价值也是不一样的。

对于古人来说，因为贸易双方需要评定商品的价值，必然会对不同商品在分割之后的特征进行分析，哪些商品可以任意分割、哪些商品可以适当分割、哪些商品根本不能分割，这都不能大意。这些问题

反映到一般意义的物质上，就不由得令人好奇：要是把一种物质无限分割下去，会发生什么？

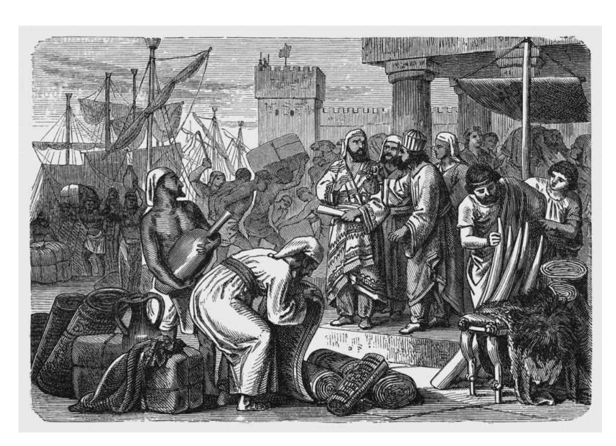
可见，先民对物质的探索，就不只是一件吃饱了没事干才会去想的问题。从古至今，几乎所有的哲学家都会表述自己对物质的看法。

## 古希腊先哲们在思考

大约距今3 000年的时候，位于现代希腊到土耳其一带的那片地方纷纷建起了一些城邦，像雅典、斯巴达之类的地名，都起源于那个时代的城邦名。

古希腊的这些城邦，即便不是地中海的港口城市，到港口的距离也不会太远。于是，他们发展出精湛的航海技术，在各个港口之间穿梭，并形成了发达的商贸文化。来自天南海北的各种商品，都在这里交汇，形成了一张庞大的网络，贸易复杂程度远不是买牛犊和木犁这么简单。

正如前面所说，商品的分割成为迫在眉睫的论题，对它刨根问底，就是在寻找物质的本源，一大批思想家对此争论不休。



古希腊城邦经商盛景

最初，在一座叫米利都（Miletos）的古希腊城邦（今属土耳其），涌现出大批哲学家。米利都的旧址位于现在的小亚细亚半岛西

岸，在陆地上只能算是偏僻的边陲地区，但是从海上来看，却是一座四通八达的交通枢纽。在公元前8世纪后，米利都成为古希腊工商业和文化中心之一，也是在东部最大的城市。当时，除了希腊各城邦外，无论波斯、新巴比伦还是埃及、叙利亚，贸易路线都离不开米利都。可以说，米利都占尽地利，也因此汇集了各地的思想。

关于物质的本源，在如今留下的记载中，最早就来自米利都的泰勒斯（Thales，约公元前624年—约前547年），他甚至被称为古希腊的第一位哲学家，被誉为“希腊七贤”之一。他的思想影响深远，甚至由此诞生了以他为首的古希腊第一个哲学学派——米利都学派。由于米利都所处的地区在当时被称为伊奥尼亚（Ionia，也叫爱奥尼亚），所以他建立的这个学派的理论成为伊奥尼亚哲学的主要组成部分。

泰勒斯原本的兴趣是在几何学与天文学的研究上，提出过一些几何定理，为此还学习了埃及和巴比伦的相关知识。他最为后世所称道的，就是曾经成功预测了公元前585年5月28日这一天的日食。虽然这个典故至今还存有疑问，但他的确利用自己的所学，估算出了太阳的直径，并解释了日食的形成原因。对于航海贸易而言，海潮、风向都可能会让一笔原本获利颇丰的买卖瞬间打了水漂，因此，各种天文现象都很重要。可以想见，像泰勒斯这样精通天文学的哲学家，在当时商人们的眼中就如同神明一般，人们必然也会将“物质本源”的答案寄托在他身上。

他说，物质的本源，就是水。万物皆由水而生成，又复归于水。

这个离奇的想法，实在有些让人难以理解。对此，他解释道，水本身是液体，可以结冰，也可以变成气体，而自然界中所有的物质，无非就是固、液、气三种状态。可见，水是万物之源，万物都有水的特性，水的特性也在世界万物中都体现了出来。他还说，连大地都来源于水，就像埃及泛滥的尼罗河会把淤泥冲积成滩涂或三角洲一样，我们生活的这个大地也漂浮在水上。

显然，他并没能解释明白物质的本源，甚至回避了实体的物质分割后是否还是原来物质的基本问题。

但是不管怎么说，这是人类历史上第一次有人清晰地阐述了万物

之根本，泰勒斯思考的精神会一直留下来。

在此之后，米利都学派的其他成员又进一步发展和修正了泰勒斯的观点。有人提出了气，有人提出了火，还有人提出了土。最后，出生于阿克拉加斯（今意大利阿格里根斯）的古希腊哲学家和诗人恩培多克勒（Empedocles，约公元前495年—约前435年）提出四元素说（即“四根说”）。他把万物的本源称作“根”，认为水、气、火、土就是这个世界上的四种基本元素，它们是不变和永恒的，不能自己运动和相互转化，但可以由这四种元素按不同比例组合和排列，构成不同性质的物质。换句话说，把世界万物进行切割，最终就会得到不同比例的这四种元素。这些元素都是以我们看不到的微粒形式存在，它们之间通过作用力结合在了一起。

恩培多克勒创造性地提出了元素的假设，经过这样的修正之后，物质本源的问题就有了一个还凑合的答案。如果万物的本质都是水，那我们不太好解释，为什么单一的水可以形成这么多种物质。但是，如果万物的本质有两个元素，解释起来就容易多了。就像水和面粉混在一起那样，水多的时候是糨糊，水少的时候是面饼，可以制造出许多种面食来。现在，世界万物的本质有四个元素，它们之间的比例可以是千变万化，通过调和，构造出世间万物，出现这样的结果并不意外。

恩培多克勒的这个想法，某种意义上说还真是没有说错。在第7章里，我们还会看到现代科学理论中，仅仅依靠四种基本的化学元素，就构造出各种神奇的生命分子，这正是恩培多克勒所设想的模式。

但是，直到这个时候，要想说明白物质被彻底分割后的本源究竟是什么，依然还欠点火候。我们并不知道，恩培多克勒所说的水、气、火、土，是否就等同于客观存在的这些物质。在他自己的理论框架中，与其说四种元素是一个个真实存在的物质，还不如说是意识在物质世界中的反映，四种元素理论说成为一种象征，就像中国的五行学说一样。

五行的本意，代表的是金、木、水、火、土五大行星。用这些词

汇来对这五个行星命名，很难不让人联想到恩培多克勒的“四元素说”，因此，五行也被称为中国古代思想家的“五元素”学说。他们把金、木、水、火、土五种物质作为构成万物的元素，以说明世界万物的起源和多样性的统一。虽然它们是真实存在的物质，但它们代表的内涵，却复杂得超乎想象，而且，五行之间还有紧密的联系，形成相生相克的关系，具有朴素的唯物论和自发的辩证法因素。虽然“五行”说后来被唯心主义思想家神秘化，比如人体内的心、肝、脾、肺、肾这五种器官，也被纳入五行之中，心属于火，肝属于木……尽管我们并不能从心脏中看到火苗，更不会看到肝发芽结出种子，但有关五行的很多合理性解释还是被保留下来了。“五行”说对中国古代天文、历数乃至医学等的发展起到了一定作用。




相生相克的五行

同样，在恩培多克勒看来，四种元素也有虚拟的一面，这种朴素的唯物主义学说可由希腊神话中的四神代表：宙斯是火，赫拉是气，涅司蒂是水，而埃多涅乌是土。这样的象征意义，早已渗透在各种文化之中，我们直到现在也还可以看到。比如黄道星座，除了蛇夫座以外的十二星座，在占星学上具有重要地位，它们就是按照四种元素被分为四象，循环往复。这里的四元素就和五行一样，和实际的物质已经没有多少联系了。

就在恩培多克勒降生前不久，米利都还迎来了另外一位哲学家留基伯（Leucippus，约公元前500年—约前440年）。关于他的历史记载并不是很多，后人猜测他成年后的主要活动地区是在色雷斯

（Therace）城邦，在那里他结识了另一位哲学家德谟克利特（Democritus，公元前460年—前370年），并以师生相称，将平生所学倾囊教授给他的这位学生。

留基伯的观点可能也受到恩培多克勒的启发，认为物质是由很多微粒构成，只不过他眼中的微粒并不是那么虚无缥缈，而是真实存在的。而且，不只是有水、火、气、土的微粒，万物都有各自的微粒，它们的本质相同，但是大小、形状以及运动的方式不同。

德谟克利特进一步发展了留基伯的理论，给这种微粒起了一个名字，叫a-tom，也就是后来的atom（原子），从而形成了欧洲最早的朴素唯物主义的原子论。他们认为，宇宙万物是由最微小、坚硬、不可入、不可分的物质粒子——原子所构成的。他把恩培多克勒的元素学说也融合进来，认为原子没有那么多种类，而是分别隶属于水、火、气、土这四种元素。每一种原子在性质上相同，但都有各自的形状特点，其大小是多种多样的。我们既不能将它们分割，也不能创造出它们。按照不同的形式组合，就可以构成所有的物质。德谟克利特的原子论可以解释日月、星辰以及天体形成的原因，甚至其认为人的灵魂也是由原子构成的。

德谟克利特的原子论是难能得可贵的，虽然在当时的条件下，无法得到科学实验的验证，但却能被人们所接受。至此，人类终于猜测出物质的本源——并用“原子”为它命名。尽管这个理论后来被证明仍然存在很多错误，但它构建的模型却与2 000多年后的科学理论不谋而合。到19世纪初，这种学说在新的历史条件下逐步发展成为近代的科学原子论。

## 给原子排排队

德谟克利特的猜想，曾经被古希腊大哲学家、思想家柏拉图（Plato，公元前427年—前347年）采纳了一部分。此时，希腊哲学的中心已经从米利都转移到了雅典，柏拉图正是雅典学派的代表人物之一，柏拉图还把他的学问传授给了学生亚里士多德（Aristotle，公元前384年—前322年）。但是，亚里士多德并不很相信原子是真实

存在的，转而研究起恩培多克勒的想法。在他看来，有没有原子并不重要，只要元素的性质经过调和，就可以形成千变万化的物质。

亚里士多德是古希腊哲学家，其影响力巨大，在多个科学领域的发展都做出了很大的贡献。在哲学上，他提出潜能与实体说，解释了世界的运动性和变化性，但是他对原子的漠视，也让后世的很多人都不再认为物质是由一个个真实存在的小微粒构成——等到人们意识到这是个错误时，已经是17世纪的事了。此时德谟克利特的名字都快被人们遗忘，更别提原子的假说了。当英国科学家罗伯特·波义耳（Robert Boyle，1627年—1691年）和艾萨克·牛顿（Isaac Newton，1643年—1727年）这样的大科学家都在猜想物质中的微粒时，他们都没有想起使用“原子”这个词。

直到1808年，英国科学家约翰·道尔顿（John Dalton，1766年—1844年）才又正式启用了“原子”的概念，发表“原子学说”，首次提出物质是由不连续的最小微粒——原子组成的。他将原子视为构成物质的最小单元，合理地解释了当时已经发现的化学现象。不同于德谟克利特，道尔顿并不只是从逻辑上猜测原子的存在，而是根据当时已有的实验结果证实原子存在。他同时也对原子设定了几个规矩：元素最基本的粒子就是原子，不可分割，在化学变化中保持不变；同一种元素的原子，形状、质量和性质都相同；不同元素的原子能够以自然数的比例相结合。

道尔顿的这些论断就是现代科学认识理解原子的基础，奠定了近代化学的科学理论基石。尽管原子很小，但是我们不能因此就否定它们的存在。还有一点是，道尔顿所说的元素，也早就不再是水、火、土、气这四种凭空猜测的元素，而是此前由波义耳提出的约定——无论怎样操作都不会被分解的单一物质，如氢、氧、碳、铁等。

显然，这里的“元素”，其实更应该被称为“单质”——由同一种元素的原子构成的物质，只不过，当时的人们并不知道原子还会构成分子，误认为所有的单质都是由一个个原子直接堆积而成，所以，单质自然就被视为元素本来的面貌。其实，在道尔顿那个时候，也已经有少量的证据对这种论述提出了质疑，比如同样是仅由碳形成的物质，既可以是石墨，又可以是钻石，那么到底哪一种才能代表碳元素呢？



神奇的碳

后来，在此基础上，元素的内涵得到了修正，它成为一类物质的总称。就好像“猫”这种动物有很多血统，可以是黑狸花，也可以是波斯猫，每个血统都不能代表整个物种。在这里，单质好比是纯种猫，而元素就好比是物种，至于原子，指的当然就是个体了。再后来，元素在化学上发展成为不能再分解成更简单的物质的概念，就是我们现在所说的化学元素的简称。

有了这样的区分，我们终于可以明白，亚里士多德那个年代对元素和原子的争论，多少有些盲人摸象。原子是构成物质的真实个体，而元素是对不同原子的分类，它们不过是描述物质的一体两面。

但是，既然用元素对物质进行分类，那么这个地球上到底有多少种元素呢？19世纪的很多化学家都在研究这个题目，而他们的依据，就是道尔顿对于不同原子的论断——相同的元素有着相同的原子，那么如果原子不同，大概就是不同的元素吧？

虽然原子太小，肉眼无法看到，但是科学家们却有很多办法识别出原子是否相同，其中最重要的一条就是测算不同原子的质量。

就这样，道尔顿提出原子论的时候，人们还只能胡乱地猜测出十几种元素。半个世纪过去以后，人们却已经可以准确地识别出五六十种元素。

这么多种元素，想要记住它们也不容易。于是，有些科学家就想了个办法，把不同的元素按照一定的顺序排列起来，最简单的依据自然还是原子质量了。

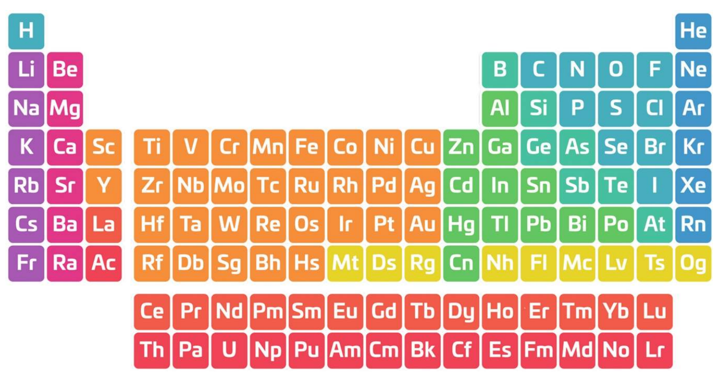
这一排不要紧，有人发现，不同的元素之间好像还有着某种规律：按照原子质量从小到大的顺序，似乎每隔几个元素，它们的性质就会轮换一个周期。就好像我们编排日历的时候，每隔7天，就会依次从星期一排到星期日。

元素的这个规律，当时很多人都认为只是一种巧合，但也有几位学者认为，这些元素的背后，应该还藏着某种未被发现的神秘力量。

到了1869年前后的时候，俄国科学家德米特里·伊万诺维奇·门捷列夫（Дми́трий Ива́нович Менделе́ев，1834年—1907年）总结了前人发现的各种现象和规律，正式提出了“化学元素周期律”，并据此绘制出著名的元素周期表；表中各元素是按原子序数由小到大依次排列，元素的性质随着原子序数的增加而呈周期性的变化。虽然他说的有一些道理，但是质疑他的人很多，因此，元素周期律一开始也就没有引起什么反响。

新生事物的诞生往往不是一帆风顺的，甚至会受到非难或指责。不过，门捷列夫似乎已经预判了这个局面，在表述时还留了个心眼。在他的元素周期表中，他特地留了4个空格，预言了一些尚未发现的元素，声称这些位置将会有新的元素填充进去。1871年，他又把预言元素的空格由4个改为6个，并且把这些元素的性质都给预测了。元素周期表为寻找新元素提供了一个理论上的向导。

这个办法就好像是在做数列游戏。比如，如果给我们一列数字：1，1，2，3，5，8，13，（？），34，那我们很容易猜到，括号中问号的数应该是21，这是一段非常出名的斐波那契数列（或斐波那奇数列）。门捷列夫的预言也是这样的原理，只是需要等候一个在括号中填写数字的人，一旦结果契合，自然也就印证了他的论断。



现代元素周期表

仅仅经过10年，门捷列夫预测的新元素应验了。例如，1875年，原子量为68的“类铝”（符号为Ea，意为类似铝的某元素）被发现了，被命名为镓（Ga，原子量69.7）。原子量为45的未知元素——“类硼”（符号为Eb，取名为ekaboron，意为类似硼的某元素）于1879年被瑞典化学家拉斯·弗雷德里克·尼尔森（Lars Fredrik Nilson，1840年—1899年）发现了。他用拉丁语中表示Scandinavian（斯堪的纳维亚半岛，瑞典和挪威就位于此岛上）的词语将这个新发现的元素命名为钪（Scandium，符号为Sc）。钪的相对原子质量为44.95，正是门捷列夫预测的那个缺失的元素。1886年，“类硅”（符号为Es，意为类似硅的某元素）也被发现了，被命名为锗（符号为Ge，原子量72.6）。根据当年门捷列夫关于元素周期律的猜测，这种新元素应该和空格元素在很多方面非常相似，事实上也的确如此，发现者都佩服得五体投地。有了这样的验证以后，科学界再也不能对门捷列夫的发现熟视无睹了。

至此，门捷列夫对按照原子质量顺序找规律的这个方法也非常信任，只是他始终搞不明白，为什么后来总有几个元素的顺序不太对，就好像刚刚过完星期一，时间又回到星期日了。

这一切，其实都源于他对原子最根本的执念——原子是物质不可分割的最小单元。

而在第1章的结尾，我们已经知晓，是比原子更小的一些微粒构成了原子。因此，原子不仅可以继续分割，而且相互之间还可以发生转变。

好戏，才刚刚上演。

## 永不停息的融合

门捷列夫临终前，听说了这件他最不愿意相信的事：原子是由更微小的微粒构成，其中至少存在一些带负电荷的电子，还有一个带有正电荷的原子核。

他之所以不愿相信，是因为一旦存在这种可能性，他所建立的元

素周期律，很可能就要崩塌。那是他一生中最为得意的作品，他不想就这样放弃。

然而，事实证明，门捷列夫多虑了，元素的周期律恰恰源于它更精细的内部结构，我们还会在后面谈及此事。而且，这样的新发现也不会削弱门捷列夫的历史地位，只会让人更加感到他寻找自然规律的本领不可思议——在没有发现原子的结构前，他居然只靠草稿纸上计算的数据就推断出了如此精妙的自然规律。

在所有原子中，最微小的氢原子我们已经见识过，它由一个质子还有一个电子组成，结构非常简单，质子便是它的原子核。

然而，在宇宙大爆炸后不久，质子和中子“抱”在一起的那个结合体，它的化学特性居然也和只有一个质子的氢原子非常相似，看起来属于同一种元素。电子比质子小得多，在原子质量中可以不去考虑，可是中子的质量和质子差不多，一个质子加上一个中子之后，原子的质量就翻倍了。按照门捷列夫的观点，原子质量决定元素的特性，这两种物质的重量相差一倍，它们的特性应当大相径庭，怎么还会有这样相似的结果呢？

类似的情况还有很多，因此科学家们在门捷列夫的研究基础上又开展了很多实验，终于认定：元素的性质和原子核中的质子数量有关，和中子的关系不大，和原子量之间自然也就没多少关系了。只不过，质子越多的原子核通常中子也更多，原子的质量相应也会更大。所以，对于绝大部分元素来说，门捷列夫猜测的依据都奏效了。这既是一种有些巧合的自然规律，也是我们人类的幸运——否则我们还要再等待更久的时间才能迎来那个发现元素周期律的人。1906年，诺贝尔奖委员会拟将化学奖授予门捷列夫，但遭到瑞典皇家科学院个别科学家的强烈反对。次年2月，门捷列夫与世长辞，成为诺贝尔奖史上一大遗憾。

随着天然放射性现象（1896年）和同位素（1910年）的相继发现，人类对原子结构的认识更进一步；还有人工合成元素的进展，它们又使元素周期表得到不断被丰富和发展。

不管怎么说，我们现在已经确信，当原子核中只有一个质子时，

它就属于氢元素。反过来，构成原子核的，除了这个质子以外，可以什么都没有，但也可以有一个中子。为了区分这两种氢，没有中子的一种被称为氕，有一个中子的则被称为氘。在宇宙之中，充斥着大量的氕和氘。

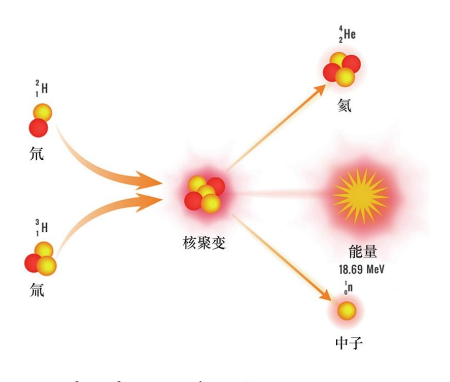
实际上，除了氕和氘以外，氢原子还有第三种形式，就是由一个质子和两个中子构成，被称为氚。氕、氘、氚在汉字中的写法，就已经形象地标明了它们的内部结构。因为它们都属于氢这一种元素，原子序数相同，在元素周期表上占据着同一个位置，只是中子的数量和质量数不同，且化学性质几乎相同，所以它们就是氢的三种“同位素”。不过，氚并不会很稳定地存在，所以宇宙中的氢原子，主要还是由氕和氘这两种同位素构成。绝大多数元素都有多种同位素。

在宇宙中出现了大批的氢原子以后，因为各种引力的关系，它们就团簇在一起，形成大片的尘埃云。此时，原子之间会发生非常复杂的相互作用，其中有一些作用，我们还会在后面的章节中了解到。

这些以氢原子为主构成的巨大云团，在蓄积到一定体量的时候，就会开始坍塌。所谓坍塌，就好比我们嚼着口香糖吹起一个泡泡，泡泡破了以后，口香糖就会立即收缩，糊在嘴唇上。只不过，造成氢云团坍塌的原因，是那个让牛顿想破了脑袋的“万有引力”，它吸引着云团外围的原子向着中央飞去。

坍塌的云团让原子之间的距离越来越小，打破了原有的平衡。紧靠在一起的氢原子会相互摩擦，产生热量，以至于它们的温度越来越高。高温会让它们再次失去捕获的电子，成为孤独的氢原子核，同时不断的挤压又会让原子核之间的距离靠得越来越近。

如果不是这样极端的环境，很难想象如此多的氢原子核会相互紧挨，它们自身携带的正电荷本应该让它们之间同性相斥。挤在一起的氢原子核靠得实在太近，它们相互撞击，终于在条件合适的时候，再一次引发了爆炸。比起宇宙大爆炸来说，这样的爆炸虽然规模小得多，却也足够绚烂——第一批恒星正是因此而被点亮。




$$^{2}_{1}H + ^{3}_{1}H \longrightarrow ^{4}_{2}He + ^{1}_{0}n + 18.69\ \text{MeV}$$

核聚变机理

在恒星内部的爆炸中，氢的原子核——主要还是氕和氚——会发生融合，这个轻原子核聚合为较重原子核的过程就被称为核聚变。聚变的同时将发出巨大的能量。它们的聚合方式十分多样，例如，一个氕和一个氚，它们就可能发生非常简单的核聚变过程，直接融合在一起，从而得到有两个质子和一个中子的原子核。有两个质子的原子属于氦元素，所以这是氦的一种同位素。只不过，氦元素不像氢那样受到特殊优待，人们并没有给它的同位素赋予各自的名称，一般就用氦-3称呼这种同位素，其中3代表的是同位素中质子与中子数之和。

这种氦-3还会和氕继续发生核聚变，产生氦-4，也就是包含两个质子和两个中子的原子核，剩余的那个质子——也就是氕核，又会接着和其他原子核相撞。将氢的同位素氕与氚的原子核无限接近，在特定条件下可使其发生聚变而形成氦核，同时放出一个中子，就能释放出巨大的能量。

宇宙大爆炸初期产生的原子核种类还有好几种，除了氢以外，还有少量的氦，以及拥有3个质子但也更微量的锂元素，它们都会参与核聚变过程中。

就这样，原子核不断地撞击聚合，形成了新原子核，总质量会有些许降低。而我们将在第5章说到的爱因斯坦质能反应方程式，在这里也一样适用，因此在核聚变过程中，会释放出巨大的能量。正是如此巨大的能量，支撑着核聚变继续发生，爆炸也在不断地进行。

正如春秋时期思想家、道家创始人老子在《道德经》中所说：“道生一，一生二，二生三，三生万物。”原子核中的质子数目逐渐增加，新的元素由此出现，万物都以此为基础。以老子为代表的东方哲学家，并没有提出和希腊哲学家那样旗帜鲜明的原子论，但是不可否认的是，他们的物质观却以原子变化的形式得到了印证。

整个过程就像是多米诺骨牌一样，一旦启动就不会停止，拉扯出产生新元素的链条：6个质子的碳、8个质子的氧、14个质子的硅……它们在核聚变的链条中占据优势，比例也较其他元素更高一些。

一般条件下，发生聚变的概率很小。自然界只有在太阳等恒星内部，因其温度极高，轻核才有足够动能克服斥力而发生持续的聚变。实现聚变反应需要上千万摄氏度以上的高温和高压。研究受控热核聚变是解决能源枯竭的重要途径。核聚变与太阳发光发热原理相同，因此，可控核聚变研究装置又被称为“人造太阳”。核聚变原理看似简单，但要让聚变反应持续可控，可以说，难于上青天。据新华社2021年5月28日报道，通过40年的努力，有“人造太阳”之称的全超导托卡马克核聚变实验装置（Experimental Advanced Superconducting Tokamak, EAST）创造新的世界纪录，成功实现可重复的1.2亿摄氏度101秒和1.6亿摄氏度20秒等离子体运行，向核聚变能源应用迈出重要的一步，未来可建设聚变电站。

不断的聚变过后，原子核越来越大，恒星中原子的数量却在不停地减少，因此恒星的内部就如同一座被挖空的山洞一般，若不是核聚变产生的巨大能量，随时可能发生坍塌。

当核聚变来到拥有26个质子的铁时，这个结果还是如约而至了。对于恒星而言，这是一场悲剧，它意味着恒星的生命就要走到尽头，但是对于整个宇宙而言，这样的灾难经常在发生，并不值得大惊小怪。

更为重要的是，它让更多元素的诞生成为可能。

当巨大的恒星坍缩之时，原子之间的碰撞也是盛况空前，灿烂的超新星也会由此形成。氧、硅、铁等有着比氢原子核大得多的原子核，它们以一种视死如归的劲头融合在一起，从而形成了那些比铁更

大的原子核。这其中，就包括铜、锌、金这些将会在我们的后续篇章中出现的元素。原子核中的质子数量飞快增长，等到出现铅元素的时候，质子的数目已经达到了惊人的82个。



一句话介绍一种元素

比铅更大的元素还会继续形成，比如铋、铀、钚等。只不过，它们的原子核实在过于庞大，已不再能够保持稳定。每过一段时间，这些原子核中就会有一部分发生分裂，变成小一些的原子核——这个过程，就被称为核裂变，它是一种与核聚变相反的过程。核裂变的存在，也注定元素的种类不会无限增加。

至此，盛极必衰，这颗恒星已经无力回天，元素的盛宴也该收场了。上百种元素交汇在一起，形成了新的云团。令人吃惊的是，在这个云团中，那些没有在核聚变中消耗完的氢原子居然还是主力，它们正打算故伎重演，至于那些因它形成的各种元素，却已经在酝酿新的物质故事——我们在下一章中继续讲述。

# 3　让原子组合起来——物质世界是如何组装的

## 宇宙分子

茫茫宇宙之中，原子所占的空间非常有限，有如茫茫戈壁滩上偶尔出现的行人，似乎很难碰撞到一起。但是，物质之间的相互作用力，却让原子上演了一出又一出的好戏，恒星就是大量氢原子和氦原子碰撞之后产生的壮丽焰火。太阳就是一颗恒星。维持恒星辐射的能源是聚变反应，即热核反应。

但是，宇宙中原子的碰撞并不总是如此激烈，核聚变产生的条件并不是那么容易达成。更多的时候，好不容易聚集到一起的原子，只是构成了非常稀薄的原子云团，没有强烈的挤压，也不会形成很高的温度，它们只是以一种更为和谐的方式聚集在一起。

在原子中，原子核是原子的核心部分，其体积只占非常小的一部分，直径只有$10^{-15}$～$10^{-14}$米（即不足100万亿分之一米或10万分之一纳米）；在一般的化学反应中，原子核是不会发生任何变化的。组成单质和化合物分子的最小微粒——原子的直径为$4\times10^{-10}$～$6\times10^{-10}$米（即不足1纳米），且其质量几乎集中于原子核。可见，原子核的直径还不及原子的万分之一。如果说原子有一只篮球那么大，那么对应的原子核不过是灰尘般大小。

原子核是由带正电荷的质子和中性的中子（二者统称为核子）组成的紧密结合体，因此，原子核带正电荷。一切原子都是由一个带正电荷的原子核和围绕它运动的若干电子组成的。当温和的条件不足以让电子与原子核发生彻底分离时，那么不同原子的原子核也就没有机会可以碰到一起，原子之间的交流，只能靠最外缘的电子牵线搭桥。

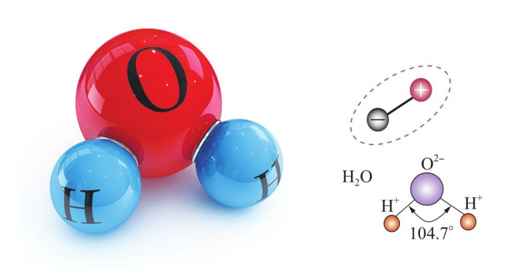
当两种或两种以上元素的原子通过电子结合，形成一个集合体，这就是原子团，而单质的分子就是由相同元素的原子结合而成的。在

许多化学反应中原子团作为一个整体参与。宇宙中的原子相遇之时，就会形成各式各样的原子团，我们姑且把它称为宇宙分子。

由于宇宙中的氢原子占据了绝大多数，因此，最容易相会的，就是散落在各处的氢原子。当两个氢原子碰到一起的时候，就会结合成氢分子，在地球上，它被称为氢气。

随着恒星内部的核聚变释放出更多类型的原子之后，宇宙分子的种类也开始多了起来。

很长时间以来，宇宙分子的神秘面纱都让人们感到捉摸不定。这是因为，在一般情况下，即使宇宙中的原子相遇了，它们周围的环境也仍然十分空旷，甚至比人类在实验室里制造出来的真空更像真空。这就意味着，原子在结合之时，很难再有别的选择，只能碰到什么就和什么结合。




水分子模型

化合物的分子则由不同元素的原子组成。比如，当氧原子和氢原子在地球上相会时，它们最容易形成的，就是供养了地球上无数生命的“水分子”。每一个水分子，都是由一个氧原子和两个氢原子结合而成，氧原子居于其中，而氢原子以特定的角度结合在氧原子的两侧，形成V形构造。考虑到氢原子和氧原子之间悬殊的体型，它们形成的这种水分子，从外形上看很像是熊猫的脑袋。

在地球上，想要从水分子上摘掉一个氢原子，让它只剩一个氢原子和一个氧原子，说容易也容易，甚至不需要太多的外力，氢原子转身就会从水分子中离开。只不过，它在离开之时，并不会带走自己原

来那颗唯一的电子，于是，剩下的一个氢和一个氧，就拥有了一个过剩的负电荷。

离子化合物没有简单的分子，是由相反电荷的离子聚集在一起的，如NaCl等。这种带有电荷的微粒被称为离子，它们的很多特性都相较于不带电荷的分子发生了变化。当微粒带有正电的时候，就被称为正离子或阳离子，脱落的那个氢缺少了一个电子，于是它就带有一个正电荷，被称为氢离子；而当微粒带有负电荷的时候，就被称为负离子或阴离子。显然当氢离子离开之后，水分子剩下那部分便是负离子。

正因为水分子中的氢很容易脱落，由一个氢和一个氧形成的这种离子，即使是在纯水中也有少量存在。由氢元素氢和氧元素组成的一价原子团就是氢氧根（$OH^-$），也被称为氢氧根离子或羟基负离子。这个“羟”字，无论是字形还是读音，都是“氢”和“氧”的“杂交体”。

从水分子上摘掉一个氢原子，还有一种特殊的情况，那就是氢在离开的时候带走了电子。也就是说，脱落的是氢原子而非氢离子，剩下的那部分，是氢和氧构成的中性微粒，不带电荷。这种微粒被称为羟基分子，它在地球上不算常见，有时候就算形成了羟基分子，通常也不能稳定存在，很快就会转化为其他物质，它们两两结合，就会变成由两个氧原子和两个氢原子组成的过氧化氢（$H_2O_2$）。过氧化氢的质量分数为3%的水溶液，俗称双氧水，也有用于漂白、杀菌作用或作为氧化剂的浓溶液，其过氧化氢的质量分数为30%。不过，双氧水依然不稳定，它在释放出一个氧原子后，就成为一个新的水分子。

可是，太空中的环境就很不一样，在地球上不能稳定存在的羟基，在太空中却有可能大量存在。实际上，早在20世纪50年代，就已经有不少人陆续推断，羟基分子是一种常见的、存在于星际空间的星际分子。

事实也果然不出所料。

1963年，科学家们首次利用射电天文望远镜通过光谱的方法，在仙后座附近探测到了羟基分子，这也成为当时一件轰动性的天文大事件，被誉为20世纪60年代天文学的四大发现之一。

人们兴奋的并不只是这种分子被证实存在，而是羟基分子和水的特殊联系，让人不禁浮想联翩——宇宙中是否也会存在水呢？多年以后，这个猜测也被证实了。

正因为宇宙的环境在地球上难以实现，各色奇怪的宇宙分子层出不穷，人们甚至专门设立了“星际化学”专业方向，用以研究宇宙中这些分子究竟是如何形成的。

继续说羟基分子。在我们知道它是星际中的常客后，很快又注意到，它并不只是会在稀薄的太空环境下出现。直到现在，我们也没有彻底调查清楚它的行踪。2021年，英国科学家首次发现，外太空一颗巨大的行星上，其大气层中居然含有羟基分子。考虑到地球也是一颗行星却难觅羟基分子的实际背景，这个结果委实让人大吃一惊。

而在2022年，中国的嫦娥项目团队在从月球上采集回来的土壤中发现，距离我们近在咫尺的月球上，居然也有羟基形式存在的水，虽然含量极低，但也还是令人眼前一亮。

所以，我们不能满足于现有的成就，还要继续探索这些宇宙分子不同的来历，这也是为了弄清楚，宇宙真实的起源究竟是什么。

实际上，很多工作早已在开展之中。尽管地球的自然环境并不满足要求，但是科学家们却在勇敢地克服各种困难。他们有的把实验搬到环绕地球的空间站中进行，有的则是在实验室中制造出特殊的条件。这些条件虽说并不像前面说过的欧洲核子研究中心那样惊天动地，但是最终的目的却是惊人的一致。

2017年，中国大连化学物理所的几位科学家，通过自制的一种光源设备让水分子发生了分解。令人兴奋的是，喜欢空手离开的氢原子，这一次却没有忘记自己的电子，于是中性的羟基分子就在这样的环境中稳定地形成了。研究人员们猜测，在宇宙中，或许有一些羟基分子，也是在类似的环境中形成。如果是这样，我们就可以通过验证羟基分子的存在，来论证那些遥远星系内的特殊状态，进一步找到宇宙形成的依据。

宇宙是物质的，通过物质去理解宇宙，这是我们永远都不会停下脚步去探索的艰巨任务。

但是，宇宙中寻找到的特殊分子，也让我们对地球上的分子有了更多的理解。如果说，原子是构成世界万物的基本单元，那么分子所扮演的角色，就是让这些基本单元发挥出实际的功能。只有认识了分子，才能真正弄明白物质搭建的规律。

## 从太空到地球

和羟基分子一样，宇宙中还有很多非常奇特的星际分子，科学家至今已经发现了其中的110多种星际分子。每一个分子都有自己的特定结构，这也是人类区分它们的依据。很多迹象表明，某些结构有可能只在太空之中才会稳定存在。

20世纪60年代，宇宙起源这个话题开始从天马行空的遐想转向以实验证据为主导的阶段，天文学家、物理学家、化学家乃至生物学家都联起手来。这样的合作很有必要：过去，理论学派的科学家们只能依靠周密的计算与必要的猜测，想象那些星际分子在宇宙中演化的过程；但是现在，实验学派着力于合成出这些分子——这甚至不只是起到辅助作用。

羟基分子已经是一个很典型的案例。多数星际分子是稳定的化合物，在地球上都可以找到；少数的星际分子在地球上很难找到，甚至根本找不到。它们有的是离子分子，在地球上虽然不能稳定存在，但在过去的实验研究中，人们早就已经知道这种物质的存在，也很熟悉它的各种光谱特征。因此，当天文学家费尽心思从宇宙中获得了相关参数后，再想确定它的存在，实际上已经不存在多少障碍了。通常认为，星际分子的存在与恒星形成早期和演化晚期有着密切联系。

然而，随着望远镜越来越先进，由此捕捉到的信号细节也更完备。如果这些信号来自于某种地球上不存在的物质，那我们又如何能够证明这一点呢？

这样的悖论其实早在19世纪时就已经开始对科学界提出挑战了。

1868年，法国天文学家皮埃尔·让桑（Pierre Janssen，1824年—1907年）在研究太阳光谱时发现，有一些谱线来自于一种未知元素，

而这种元素在地球上尚未被发现。于是，这个未知元素就被暂定名为“helium”，其含义是“太阳元素”。

然而，太阳元素究竟是什么？地球上只要找不到这种元素，这个问题就始终无法给出答案。

现在我们都知道了，所谓的太阳元素其实就是第2章所说的氦。在太阳中，它是仅次于氢的第二大元素，但是在地球上，它却稀缺得令人抓狂。直到1895年，也就是“太阳元素”最初被发现后的27年，英国科学家威廉·拉姆塞（William Ramsay，1852年—1916年）才从钇铀矿物中通过放射性元素的裂变找到了它，说明地球上也存在氦。

某种程度上说，当氦元素真切地呈现出来时，还是有些出乎意料，毕竟几十年前人们用金属元素才有的“-ium”后缀给它命名，而它和金属元素之间的关系却八竿子也打不着。

尽管这段故事曲折离奇，却让太阳系的形成过程有了更精确的答案。

太阳系很可能来自于一场超新星爆发后的残骸，从氢、氦、锂这样的轻元素到铁、铜、金这样的重元素都像摔碎的玻璃一样，以尘埃云的形态一堆一堆地分散在太空之中。超新星是爆发变星的一种，当它爆发时，会释放出无比巨大的能量，且星体中的大部分甚至全部物质被抛散。这种爆发变星具有亮度突增的特点。

作为燃料的氢和氦元素并没有被消耗太多，依然是这片尘埃云的主体。与此同时，更重的那些元素也构成了新的聚落。在引力的作用之下，这些云逐渐收缩成一个个小球体并旋转了起来，小球体又在旋转过程中不断地汇集，成为固定轨道上的大球体。

在漫长的整合之后，几乎所有的氢和氦共同组成了一颗硕大的球体，并又一次引发了核聚变，形成了我们如今所看到的太阳。那些没有来得及跟上脚步的氢、氦元素则在远方组成了诸如木星、土星、天王星以及海王星这样的气态星球。相比于太阳，它们的体量还是太小，并不足以点亮核聚变的光芒。于是，在太阳系中，便只有唯一的一颗恒星，其他星球绕着太阳旋转，行星的数量也屈指可数。

至于那些重元素，在一般条件下，它们不但发生聚变的概率非常小，而且数量实在少得可怜，只够组成一些更为娇小的行星，也就是从太阳到木星之间依次排开的水星、金星、地球和火星，它们都被称为岩石星球。当然，还有像冥王星、谷神星这样没有被列入八大行星的星球，以及月球这样绕着其他行星旋转的卫星。

严格来说，这个过程还有很多细节等着我们继续去探索，例如木星内部，其实还有一个数倍乃至数十倍于地球大小的巨大岩石核，它似乎告诉我们一个更有可能出现的早期物质世界：太阳系孕育时，各大星球的元素构成比例并没有太大区别，重元素形成岩石，而氢、氦这样的气态元素包裹于其外，只不过太阳的体积实在过于庞大，那些靠近太阳的行星以及体积太小的行星，都因为引力不足而丢失了氢和氦织就的外衣。

所有这些故事，都需要依靠物质提供的拼图去一一解开。事实上，我们不难证明地球早期的大气中含有大量的氢气，它们后来很多以水的形式留在了地球上。

在地球以外率先发现氦元素，给了人们很多启发，但也是个提醒：如果在太空中找到新物质，地球上却不能予以验证，会给科学研究带来无尽的麻烦。

到1968年时，天文学家们发现，银河中心的星云传递出很多分子的信号，虽然已有包括水分子在内的20多种分子得到了验证，但是还有很多信号，居然和以往所知的任何物质都不一样。此时的问题摆在化学家面前，他们务必尽快在地球上找到这些信号所对应的分子。

在连续攻克几大难题后，科学家们开始向光谱上217纳米波长的一处吸收峰进军，试图确认是哪种分子造就了这一现象。一开始，由碳原子构成的石墨分子成为最热门的候选对象，但是在被实验结果否定后，当时最前沿的天文学家相信，这些分子很可能是一类被称为“氰基聚炔烃”的物质。这类物质的主体仍然是碳原子，只不过碳原子破天荒地以直线的方式相连，两端分别是一个氢原子和一个氮原子。

于是，对星际分子的研究，就从太空“搬”回了地球上。在投入无

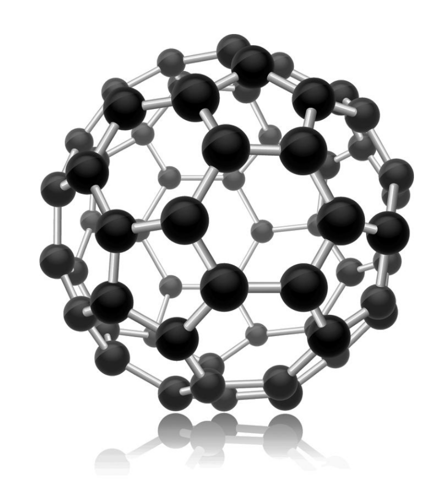
数人力物力后，这个问题至今还是悬而未决。

这似乎也在告诉我们，关于物质，深邃的宇宙中还有很多未解之谜，我们的探索生涯永远都不会停息。

只要是探索，就不会无功而返。就在寻找“氰基聚炔烃”这种物质的过程中，很多研究团队都积极开发了新的合成方法，这些方法都成为科学研究中的无价瑰宝。

英国萨塞克斯的哈罗德·克罗特（Harold Kroto，1939年—2016年）是一位合成领域的实验大师，他也于1975年加入了这场科学盛宴。他曾经合成出越来越长的氰基聚炔烃，通过对这些分子的参数进行计算，眼看着距离217纳米的目标越来越近。然而，地球的环境还是未能给他带来好运，当碳原子的数量超过9个时，他的合成手段已经失效。

但他设计的第二代仪器，却意外地获得了碳原子数量为60的一种分子，这让他吃惊不已。经过数百次重复实验后，他和很多同行都已经明白，这是一种过去从未发现的物质结构，60个碳以一种近乎完美的对称形式组合在一起。



$C_{60}$结构示意图

在确认这种物质结构时，克罗特想到了一位名叫巴克明斯特·富勒（Buckminster Fuller，1895年—1983年）的美国建筑设计师。富勒

曾在蒙特利尔世博会时用六边形为主的结构搭建出巨大的穹顶——球形薄壳建筑结构。这一非凡的创意最终给了研究团队以灵感。

以克罗特为代表的科学家们用60个碳原子的模型，以六边形和五边形交替的形式拼出了一种被称为碳-60（$C_{60}$）的球状分子——直到完工的时候，一颗足球一般的分子模型摆在他们面前。这个分子由20个正六边形与12个正五边形组成，和经典的足球缝制方式完全相同，堪称迄今为止对称程度最高的一种分子，故$C_{60}$别称“足球烯”。后来，$C_{60}$这种分子由其晶体结构分析所证实，与它同类的这些分子则被统称为富勒烯。可见，富勒烯就是碳元素的一种同素异形体，即同种元素的物质而具有不同结构，但只限于单质，如碳的金刚石、石墨、富勒烯等。

1985年，这项工作被公布出来的时候，全世界都为这种美丽的分子感到不可思议，更令人感到不可思议的是，这种绝妙的分子结构在自然界居然从未被发现！

不过，很快又有了新的证据证明，哪怕只是蜡烛燃烧产生的炭灰中，也能找到$C_{60}$的身影。我们真正应该反思的是，为什么直到现在才找到它？

或许，这就是太空给我们的提示吧。

1996年，克罗特因为发现富勒烯而分享了当年的诺贝尔化学奖。但是，我们知道，对分子的探索才只是进入热身阶段，我们还需要更深入地了解它们。

## 分子的游戏

很多时候，我们会好奇，到底是因为什么，我们的地球才会变得如此美丽动人？

要是不带任何感情地回答这个问题，那么答案就是——分子。

地球形成之初，丰富的氧元素与硅元素就已经迫不及待结合在一起，以二氧化硅的形式构成了地球的主体，如今我们称之为岩石。然

而，由于此时的地球还处在极度的高温之下，即便是岩石也因此被烤化，整个地球就如同一块巨大的熔岩球。于是，那些比岩石更重的金属沉到了地球的中心，成为如今被称为地核的结构体。而在地球的表面，熔岩还在肆虐，炽热的气体不断产生，与原始的氢气等物质一起，构成了地球的大气层。

当我们抬头瞩目明亮的金星和赤红的火星，或是在望远镜里遥望木星的大红斑时，不由会觉得，它们美得简直让人窒息。然而，若我们有一天能够走近这些行星，一定会有完全相反的感受。在这些行星的表面，只有恶劣的大气环境和贫瘠的地面——甚至在木星这样的气态行星上，我们似乎连大气和地面的分界线都找不到。

单调，是这些行星的共同特征。实际上，我们有理由相信，即便是在太阳系外，绝大多数行星也都会是相似的模样。

早期的地球大概也是如此，如今却到处都是生机勃勃的景象，其中也包括我们人类在此繁衍生息。换言之，各种各样的生灵，都依赖地球独特的环境生存。

面对此情此景，人们猜测，地球最初就好比一个大熔炉，不同的物质在合适的介质中不断碰撞，终于产生丰富的分子种类，提供生命起源最初的原料。

而在这个过程中，最重要的介质就是水。

水是一种神奇的物质，以至于泰勒斯最初将它作为物质的唯一起源。泰勒斯并没有错得很离谱，对于地球而言，很多分子的起源确实有赖于水。

通常物质存在三种状态：固态、液态和气态。几乎没有任何竞争的选项，人类以水作为标准物质设置了最通用的温度标定方式——在摄氏度的规定中，以标准大气压下水的凝固点为0摄氏度，同时以水沸腾时的温度为100摄氏度，平均切分100份，就可以得到每一摄氏度，这种定标的方法叫作摄氏温标。不过，在科学上，温度存在理论上的最低值，大约是零下273.15摄氏度，如果以此为零点进行规定，便是绝对零度。以水的三相点温度273.16开（即0.01摄氏度）规定为零点建立的热力学温标是一种不依赖于任何物质的特性的最基本的理

想温标。尽管如此，热力学温度（单位为开尔文，简称开）的每一个区间，和通用的摄氏度区间并无差别，但在生活中，使用摄氏度的场景显然更多。

以水温划分温度并不意外，因为它是地球上最常见的液体物质，并且我们也很容易看到它的气态或固态形式。相比之下，如果我们想要看到铜熔化为液态的铜水，就需要加热到1 084.62摄氏度。这是一个非常高的温度，至少对于古人来说，仅仅靠燃烧木柴实在难以企及，这也就不难理解，历史上人类为什么不能很容易地掌握炼铜技术。



在没有温度计的年代，古人是如何测温的？

更具特色的是，和类似的物质相比，水在地球上保持液态的区间实在是大得出奇。

比如地球上另一种常见的分子二氧化碳，它在气温低于零下78摄氏度时会成为固态，固态的二氧化碳为白色，形似冰雪，被称为干冰。而当外界温度高于这个温度数值的时候，它甚至不会先熔化变成液态，而是直接气化变成气态，成为我们空气中普遍存在的二氧化碳气体。也就是说，液态二氧化碳在地球上存在的温度区间是零，只有改变气压，才有可能制造出它。例如，对二氧化碳施加很大的压力，就能够形成液态二氧化碳，要是任由液态二氧化碳膨胀，又可能会制得干冰。

二氧化碳或许是个极端的例子，但是其他一些分子，如甲烷分子（$CH_4$），它由一个碳原子与四个氢原子构成。它从固态变成液态再到气态，只有21摄氏度的区间；氨分子由一个氮原子与三个氢原子构成，它的区间是44摄氏度；二氧化硫中有两个氧原子和一个硫原子，它的区间是56摄氏度……这些分子的元素组成都很简单，它们和水还有二氧化碳一样，都是地球形成初期的大气层中就已经存在的物质。

可见，在太阳系形成初期，地表上流淌的这些初始原料中，水分子维持液态的能力最强。得益于地球与太阳之间恰当的距离，这颗星球表面大部分地区的温度，在大部分时间里都可以保持在0~100摄氏度。这也就意味着，地球上可以出现很大体量的水世界，它们不断地融合交汇，形成大大小小的系统——如今我们称之为江河湖海。

尽管现代科学还不能完美地解释生命的起源过程，但是液态水的存在和能量的供给，毫无疑问是最重要的两大基础要素。

直到地球上出现生命以后，才有了更多液态区间很大的物质——乙醇，也就是酒精，它的液态区间将近200摄氏度，至于各类植物油，甚至普遍可以超过200摄氏度。

只不过，如果没有最初的液态水，又何来生命，何来乙醇或油脂这样的分子呢？

如今，当我们从水龙头下接上一碗水时，或许并不会在意这碗水中的水分子，更难得去猜测除了水分子以外还有些什么物质。然而，这碗平平无奇的水，还有其中所谓的杂质，却书写了极不寻常的物质演化史。

严格来说，我们现在看到的这些水分子，和46亿年前地球刚刚形成时的那些分子并不是同一批，但它们却有着千丝万缕的联系。

大量的水分子聚集在一起，它们就会玩起丢沙包的游戏——沙包便是水分子中的氢原子。液态水中的两个分子靠得很近时，它们就会交换各自的氢原子，速度快到令人目不暇接。

实际上，很多时候，在一碗水中随便指定一个氢原子，我们甚至很难确定它到底属于周围的哪一个水分子。正是因为氢原子处于不断交换的状态，水分子才有了异乎寻常的活跃属性。无论处于多么平静的水面之下，水分子之间都如同一群剑拔弩张的仇敌，在不停地抢夺氢原子，它们将分子的游戏推向高潮。

当它们流经岩石之时，活跃的水分子会萃取出其中的矿物质，包括钠、钾、钙、镁以及氯、磷在内的各种元素离开岩石，转而在水中富集。这个过程直到今天也没有停歇，雨水冲刷着世界各地的山体和

土壤，然后带着这些矿物质，一路奔流到海，于是海水中的矿物质就越来越多。

不只是岩石，地球早期大气层中的成分同样也会被水吸收，氨气与水的亲和力惊人，海水中因此拥有了大量的氮元素。不断喷发的火山不断释放出二氧化硫与二氧化碳，又为海水提供了丰富的硫元素和碳元素。

总之，当水覆盖地表大部分面积之时，它其实早已成为“浓汤”，其中混合了各式各样的元素，其复杂程度远甚于我们从水龙头下接的这碗水。

地球诞生之初的这锅浓汤里，可以熬出越来越复杂的物质。另一方面，包括小行星和彗星在内的天外来客们也像调料包一样，朝着地球这口锅中撒下更多的汤料。事实上，很多人还坚持认为，地球生命的源头，也许就来自这些太阳系中游荡的小天体。对此，人类也从未停止过对它们的探索，试图为生命在物质世界中的诞生找到更完整的解释。

## 不变的规则

2022年6月，日本科学家宣布，在隼鸟二号小行星探测器从“龙宫”小行星带回的岩土样品中发现了氨基酸分子。

隼鸟号系列小行星探测器是日本专门针对小行星开发的研究设备。早在2003年5月9日时，隼鸟一号就发射升空，它的目标是对一颗名为“丝川”的小行星进行探测和采样，并带回样品。实际上，这颗行星也是日本天文学家发现并命名的，被选为登陆对象并不叫人意外。隼鸟一号经历了多次磨难，于2005年11月12日在丝川上软着陆并采样，2010年6月13日成功返回地球。它从丝川小行星上带回的样品因没有受到地球上的任何污染，成为人类研究太阳系进化过程的形成的珍贵物质。

到了2014年12月3日，隼鸟二号又肩负着相同的使命，只是目标换成了引力相对更大的小行星“龙宫”，经过6年的往返，其中的回收

舱终于成功带回了5.4克岩土样品。经过仔细的分析之后，这些岩土被证实含有20余种氨基酸，其中不少氨基酸的种类是地球上已经存在的。氨基酸是一种含有氨基的有机酸，其中的α-氨基酸是组成蛋白质的基本单位，因此，人们把氨基酸誉为“生命之源”，这是首次在地球以外确认氨基酸存在的证据。

这桩新闻的出现，让很多一直持有地外行星带来生命这一观点的人们又掌握了新的证据。氨基酸是一类有机物，它对生命而言可以说是基础原料，我们还会在后面继续谈到它们。此刻，我们似乎更应该关注一件事：如果原子是构成宇宙物质的基石，那它们是按照相同的规则结合在一起的吗？

至少隼鸟探测器带给我们的答案是肯定的，这对我们来说是个积极的信号。

不妨反过来想一想：假如同样的元素，它们在地球上按照一种规律结合起来，到月球上却换成了另一套规律，等到了火星上时，差异就更大了——这样多变的物质世界，会让人类的探索变得非常困难。

实际上，正是因为我们相信物质结合的规律存在共性，才有可能足不出户就能判断数百光年外会是什么环境。就像克罗托对氰基聚炔烃锲而不舍地合成是为了验证在银河中心存在这种分子的可能性，因为一旦证实了这一结果，就可以根据这种物质的特性去推断那片区域的环境。

尽管如此，科学家们也并不是从一开始就笃定这种规律的存在，甚至为此还展开过大辩论。

19世纪初，就在道尔顿提出“原子论”之后，原子如何组合的问题就引起了很多人的好奇。这是因为，大多数物质中的元素配比似乎都有着特定的比例，比如水含有氢和氧两种元素，不管怎么转化为氢气和氧气，氢气和氧气的重量比都是1：8，而体积比都是2：1。从这些现象不难猜出，如果不同元素都是以原子这样的微粒形式存在，那么它们之间必然会以特定的方式结合在一起。当时有一些实验科学家相信这是最可能的结果，特别是法国科学家约瑟夫·路易斯·盖-吕萨克（Joseph Louis Gay-Lussac，1778年—1850年）还为此提出了一条

在等压条件下关于气体的体积随温度而变化的定律，后人称之为盖-吕萨克定律，然而道尔顿本人却不这么看。他经过计算发现，如果原子会按照比例结合，那么就可能出现半个原子的结果，这显然不符合“原子是参加化学反应最小单元”的设定。所以，他认为这种巧合不过是有些实验学家的测试不够准确造成的。

对于这样的争论，当时的很多学者都提出了自己的假设。1811年，一位名叫阿莫迪欧·阿伏伽德罗（Amedeo　Avogadro，1776年—1856年）的意大利年轻科学家发表文章提出“分子”的概念以及原子与分子的区别等重要问题。他认为原子首先会组成分子，道尔顿算出来的“半个原子”，实际上应该是“半个分子”，这种分子中有两个同样的原子，所以半个分子就是一个原子——完美地解答了道尔顿担心的问题。

然而，这种说法不仅没能迎来道尔顿本人的理解，还让更多学者感到荒谬。相同的两个原子怎么能结合在一起？这样的猜测违反了当时学术界的基本观点，而它的答案我们将在第4章中揭晓。

总之，阿伏伽德罗的“分子学说”被无情地抛弃了，但他并没有因此沮丧，而是继续完善自己的工作，为分子学说提供了更多证据。后来，他的这一学说就成了我们所熟知的阿伏伽德罗定律。

随着时间的推移，越来越多的研究者发现，原子的真实存在不容置疑，但是承认原子，必然也要承认它们特定的结合形式。在阿伏伽德罗的分子学说提出40余年后，英国科学家爱德华·弗兰克兰（Edward　Frankland，1825年—1899年）于1852年已经初步提出了我们现在称之为“化合价”的概念。化合价也称原子价，简称价，用来表示一个原子（或原子团）可以和其他原子相结合的数目，如氢是一价，所以两个氢原子和一个氧原子会结合为水分子，氧的化合价就是二价；而当它们分别和碳元素结合时，因为碳是四价，所以一个碳会和四个氢结合（即甲烷），或者一个碳和两个氧结合（即二氧化碳）。但是，他当时的概念还是比较模糊的，没有论及多原子元素彼此相结合时所遵循的原则。

实际上，到了这一步时，“分子”学说就该重见天日了。但是，当

时的主流学者却不敢推翻前人的观点，阿伏伽德罗本人也已是风烛残年，难以据理力争。

直到1860年，在第一次国际化学会议上，阿伏伽德罗的意大利老乡斯坦尼斯劳·坎尼札罗（Stanislao Cannizzaro，1826年—1910年）站了出来，通过实验加以论证，重新阐述了“分子”和“原子”的关系，将这个沉睡半个世纪的重要理论公诸于世，科学界这才恍然大悟。这一理论终于得到普遍的公认。然而，阿伏伽德罗并没有能够亲眼看到这一天，他在4年前就已经过世了。

不仅如此，坎尼札罗在后来的几十年里，一直都在践行着自己的使命，不仅彻底搞清楚分子是什么，更由此修正了过去的一些错误，完善了原子量的测定。正如第2章所说，门捷列夫编制元素周期表的依据就是原子量，他能够拥有一套完整的数据，同样也离不开这些幕后的工作。实际上，此前也有一些尝试编纂元素表的先驱，就因为原子量的数据不准确而不能自圆其说，作品最终未能成型。

坎尼札罗也擅长实验工作，他首先发现了一种化学反应的过程，至今还在被广泛应用，并以他的名字命名为“坎尼札罗反应”。在这个反应的过程中，就会出现“羟基”的身影，而坎尼札罗也是第一个提出“羟基”这种结构的科学家。

尽管19世纪的科学家们对于原子为何会结合在一起完全没有头绪，但他们走过很多弯路以后，最终还是确定了这种模式。后来，虽然分子的类型越来越多，但是没有人怀疑，有一股看不见的力让原子凑在一起，它们形成的小团体能够保持这种物质最基本的化学性质。

到了现在，通过寻找特定的分子去挖掘线索，已经成为很多领域的常规操作。不仅仅是在太空探索中如此，医生会通过寻找特定分子确定病症，刑警也会根据分子去找到犯罪的证据，这都是分子理论的实际应用。可以说，原子会组合成分子的规则，已经成为我们深刻认识物质世界的基础。

但是，它们到底是怎样结合在一起的呢？接下来，我们就来看看，物质之间有着怎样的作用力。

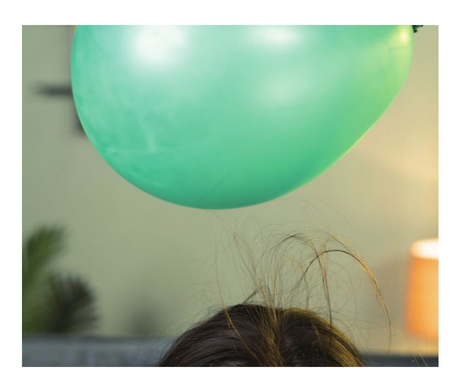
# 4　无处不在的相互作用力——物质为何能结合在一起

## 电子的巨大魔力

用丝绸在一根玻璃棒上摩擦片刻后，因为玻璃棒带上了电荷，就可以吸起一些小纸屑；同样地，把硬橡胶棒与毛皮摩擦后，硬橡胶棒也会带上电荷。物理学上把二者分别规定为正电荷和负电荷。用磁铁顺着同一个方向在铁钉上摩擦，铁钉就可以被用作指南针。

这两个经典的物理实验，讲述了宇宙间的一条重要法则——异性相吸。例如，静止的电荷，同种相斥，异种相吸。

更具体而言，带有正电荷的物质会和带有负电荷的物质相互吸引，两个磁体的磁南极和磁北极会相互吸引。也有人尝试将这个规律推演到更广泛的社会学领域，用以解释包括男女感情在内的各种问题——似乎并不总是吻合。因此，从科学的角度而言，异性相吸是在电磁学领域才成立的铁律。这些放之宇宙皆准的现象，有赖于背后的物质基础，而电子在其中扮演了最为关键的角色。



气球和头发因静电而相互作用

正如我们现在已经知道的，绝大多数物质的基本单元都是原子，

而原子的结构，是带有负电荷的电子围绕着带有正电荷的原子核旋转。原子核与电子的电性相反，使它们之间产生了一股吸引力。由于原子核的体积远远大于电子，两者之间的吸引力让它们形成一种类似于太阳系的结构：原子核如同居于核心的太阳，而电子则好比是太阳周围的行星。所以，对原子结构的这种描述方式通常也被称为“原子行星模型”。电子是绕原子核在确定的轨道上运动的，这个概念在现在的理论看来只是有限有效的，已被量子力学的概率分布概念所代替，但由于它的直观性，现仍常用轨道这个术语来近似地描述原子内部电子的运动，用作对原子结构的一种粗浅说明。

如果我们已经理解了太阳和地球之间的空间关系，构思出原子的“行星模型”似乎就是自然而然的结果，但事实远非如此简单。

根据牛顿的经典力学，宇宙的万物之间都存在万有引力，引力的大小和物体的质量以及相对距离有关，物体质量越大，或者相对距离越小，引力就越大。

然而，万有引力并不是很显著，比如一个苹果和一个橘子放在一起，它们并不会因为引力而相互靠近。只有当质量达到天体水平时，才会产生明显的效应，所以苹果和地球会相互靠近，树上的苹果成熟后便会掉落下来。牛顿的重要贡献，就是他通过缜密的数学计算证明，地球以及各大行星与太阳之间都存在着强大的引力，在引力的作用下，行星和太阳会围绕着系统的质心做圆周运动。

17世纪时，地心说和日心说的争论还在持续，牛顿提出的这些观点，在一定程度上也声援了地心说。如果两颗巨大的天体质量相仿，那么质心就位于两者的中心，当它们在万有引力的作用下做圆周运动时，更像是操场上正在进行追逐赛的两名运动员，只是谁都追不上谁。然而，太阳的质量远大于地球，在日地系统中，质心距离太阳的中心很近，所以，从远处第三者的固定视角来看，太阳几乎没有偏转，只有地球在绕着太阳旋转。

因此，若是以地球为参考，认定太阳绕着地球旋转也无可厚非，地球就是中心，“地心说”并不荒谬。但是，因为太阳与地球相对运动的本质是万有引力，而太阳产生的引力作用远大于地球，这样来看，

地球绕着太阳转，显然是更合理的观点。

正电荷与负电荷之间的吸引力和万有引力相仿，它的大小取决于带电体电荷的大小以及电荷之间的距离，电荷数值越高，或者电荷之间的距离越小，那么带电物体之间的吸引力就越强。与万有引力不同的是，电荷之间的吸引力非常显著，哪怕只是很小的带电物体，也会产生很强的作用力，这也是玻璃棒可以吸起纸片的原因。电相互作用力取决于电荷，它可以是引力或斥力；而万有引力取决于质量，它总是相互吸引的，因为没有负质量的物体。毫无疑问，这时候玻璃棒施加给纸片的电荷吸引力要显著大于地球施加的万有引力。

同样的现象在磁体中并没有能够完全对应——磁单极子至今尚未被发现。也就是说，任何一块带有磁性的物质，它都是既有南极又有北极，可能并不存在只有南极或只有北极的物质。

尽管如此，电和磁之间还是有着非常密切的联系：当磁体形成磁场时，在磁场中运动的导体棒切割磁场中的磁感应线，导体回路中的电流便形成了；反之，给螺线管线圈通电，螺旋管圈内放置的铁棒也会变成像磁铁一样。这些现象，如今早已应用在包括发电机、电动机、电磁铁等各种场景中。

电和磁之间可以相互转化的特点，早在19世纪就已经吸引了很多科学家关注，特别是詹姆斯·克拉克·麦克斯韦（James Clerk Maxwell，1831年—1879年）在1873年发表了自己的著作《电磁通伦》（*A Treatise on Electricity and Magnetism*），从理论上将这两种现象统一起来，也由此奠定了现代电磁学的基础。

在麦克斯韦的理论体系中，最为人津津乐道的便是“麦克斯韦方程组”。1864年，麦克斯韦在总结电磁现象的基本实验定律——库仑定律与高斯定理、毕奥-萨伐尔定律与安培环路定律、法拉第电磁感应定律等，以及引入位移电流的概念基础上，首先将这些规律归纳为一组看起来有些复杂的偏微分方程。他不仅解释了电与磁之间的完美关系，更进一步提出了电磁波的存在——这是由电场与磁场相互作用形成的一种波。电磁波在自然界中广泛存在，任何一种高于绝对零度的物体都会辐射出电磁波，科学家们对于这一现象的研究，将在数十年

后引发一场有关物质的大讨论，我们随后就会看到。同样让人感到好奇的推论还有，电磁波的运动速度和光一致，麦克斯韦也因此确信，光实际上就是一种电磁波。毫无疑问，麦克斯韦电磁理论的建立是19世纪物理学发展史上一个重要的里程碑。

电磁波以交变的电场和磁场通过能量转换的形式在空间中以光速传播。存在于空间区域的电磁场，电场和磁场既相互依存又相互作用，随时间不断变化，因此，这种“场”是一种特殊物质。说它特殊，是因为我们不能凭感觉器官直接感受其存在，而它间接地表现出来的物质属性，包括能量、动量和质量等，具有不依赖于人的意识而存在的客观事实。或者说，包括电磁场在内的各种场是物质存在的两种基本形态之一。另一种物质存在的形式为实物，实物具有静止的质量，与场既有区别又有联系，并可相互转化。由于场与粒子有不可分割的联系，一切相互作用都可归结为有关场之间的相互作用。按照这种观点，场和实物并没有严格的区别。

尽管电磁学的定量关系已被揭示，但是它们究竟从何而来，又因何会相互关联，却仍然毫无头绪。

在麦克斯韦研究电磁学的同一时期，对各种物质施加电压，早已是一种常用的研究手段，很多时候这样操作会改变化学反应的进程，从而产生新的物质。有科学家发现，如果在一根玻璃管中充入非常稀薄的气体，压强接近于真空，然后再对气体施加电压，这时候，阴极（负极）有可能会产生一种射线。这种阴极射线，也着实令人困扰，没有人能够说明它究竟是什么。

这一切难题，都在19世纪末见到了曙光。

在麦克斯韦电磁学理论的指导下，越来越多的科学家开始熟练地掌握电磁学手段进行实验操作。1897年，约瑟夫·约翰·汤姆孙（Joseph John Thomson，1856年—1940年）在电磁场下研究起阴极射线来。和当时其他一些人的观点不同，汤姆孙引入了电磁场的装置，他经过细致的实验证明，阴极射线是一种带电的粒子流，并根据实验参数推算出了这种粒子的比荷（即单位质量的电荷）。

通过这个实验，汤姆孙最终证实，这种微观粒子所带的电为负电

荷，并将这种粒子称为电子。后来的实验表明，微观粒子所带的电荷是量子化的，即在自然界中，电荷总是以一个基本单元的整数倍出现，这个特性叫做电荷的量子性。电荷的基本单元就是一个电子所带电荷量的绝对值，称为元电荷，用*e*表示。1910年—1917年，美国物理学家罗伯特·安德鲁·密立根（Robert Andrews Millikan，1868年—1953年）应用油滴实验方法，精确地测量元电荷*e*值，证明电荷量子性，获1923年度诺贝尔物理学奖。元电荷*e*的测定，为电子论的建立提供直接的实验基础。

电学现象和电子有着直接的关系。比如说，当我们在宏观层面上观察到玻璃棒和纸片相互吸引的现象，实际上就是因为在微观层面上，这种粒子发生了转移。汤姆孙进一步推测，电子来自于物质的原子内。这个观点在当时多少有些离经叛道，因为原子最初的定义就是“不可分割”的最小微粒。与此同时，因为元素周期律而声名大噪的门捷列夫也坚信原子是物质最小的单元，这也让学术界的这场辩论更加热烈。电子的发现打破了原子不可分的经典的物质观，推开了微观世界的大门。

汤姆孙成了这场论战的赢家，电子的确是原子的一部分，很多化学反应的原理由此被揭示，我们还将在后面继续讲述。现在的问题是，电和磁之间的联系又是怎么形成的呢？

又经过20余年的探索，在对电子运动状态的研究过程中，有人猜测，电子可能接近于带电的球状粒子（或陀螺），这只是一种直观的图像。构成物质的原子、分子中每一个电子都同时参与两种运动：核外电子绕原子核的轨道运动，电子本身的自旋运动。电子在运动的同时，自身也会发生旋转，也就是自旋。自旋有顺时针，也有逆时针。当电子发生自旋时，它就成了一个有磁性的小粒子，用一个圆电流回路来等效，那么它同时在其周围产生磁效应，就像滑冰场上的舞者在高速旋转时也会在身边产生气流一般。如果一群电子的自旋方向相同，那么它们产生的磁场就会得到加强。反之，如果电子自旋随机发生，相反方向自旋的电子就会抵消各自的磁场，磁场就会被削弱。这个等效的圆电流叫做分子电流。或者说，分子电流是分子或原子中自由电子运动所形成的电流。分子电流假说由法国物理学家安德烈·玛丽

·安培（André-Marie  Ampère，1775年—1836年）首先提出，因此，分子电流也称安培分子电流或安培电流。

尽管这样的模型在后来更为成熟的理论体系中被证明相当粗糙，甚至还有很严重的错误，比如电子被视为球形带电粒子就存在争议。但是不管怎么说，自旋是许多微观粒子和原子核的属性之一。电子自旋这种现象已被实验证明存在，相当于其固有的角动量，而它也的确是影响磁场的根本原因。正像我们不能用轨道概念来描述电子在原子核周围的运动一样，也不能把经典的带电小球的自旋图像硬套在电子的自旋上。例如，要理解原子中的电子，进一步说明原子光谱的某些特征，还需要一个自旋量子数等，这是量子物理学的理论部分。

自此，我们不难理解，日常生活中的各种电磁现象，它们的真实载体就是每一个原子中都存在的电子。当小小的电子团结在一起时，产生的电磁作用力可以大到惊人。就说暴风雨袭来时夹杂的闪电，实际上就是因为云层在翻滚时，电子发生了迁移——电子减少的区域带有正电，而电子增加的区域则带有负电，当它们累积到一定程度时，巨大的电压又会让电子一瞬间回到原位，释放出大量的电能，并引发闪电周围巨大的磁场变化。

实际上，对于物质世界而言，电磁作用力就像万有引力一样普遍，它不只是表现在我们看得到的这些现象，更表现在每个原子的内部。

## 原子的结构

在揭晓电磁力如何在原子层面上发挥作用之前，首要的难题是要搞清楚原子究竟是怎样的结构。否则，如果我们无法确认原子内以及原子之间的电荷分布，自然也就无法分析这些电荷之间如何关联。

汤姆孙在发现电子之后，次年就提出了一种想象中的原子结构，史称汤姆孙模型，也叫梅子布丁模型或枣糕模型。据说，有一天汤姆孙在吃早餐的时候，还在思索着原子的结构问题，突然看到了餐桌上的梅子布丁，一大块布丁上嵌着一些梅子，深受启发，于是提出了一

种可能性：原子就是一个带有正电荷的大球镶嵌了一些负电荷的小电子。

这个故事一看就是牛顿被苹果砸到以后想到万有引力的翻版，但它说的多少也有几分道理，尽管当时还没有任何实验可以证明这个想法，却也很快就被科学界所接受。

当然，人们之所以会认可，除了科学方面的原因，也有一部分原因是汤姆孙当时在科学界的地位。

在英国剑桥大学，有一座非常了不起的实验室，由大科学家卡文迪许的家族亲人于1871年捐助建立，从建立至今一直都是物理学的圣殿，尤其在揭示物质世界奥秘这方面做出了不可磨灭的贡献，并在百年的时间内产生了20多位诺贝尔奖得主。麦克斯韦是该实验室的创建者。汤姆孙年少成名，不满30岁就担任了这座实验室的主任。所以，当汤姆孙发现电子并提出原子模型之时，差不多可以说，当时全世界没有任何一个人比他更懂原子。

身处科学高地的他也广纳贤才，其中有一位年轻人更是从远在南半球的新西兰慕名来到他的实验室担任助手。这位助手名叫欧内斯特·卢瑟福（Ernest Rutherford，1871年—1937年），汤姆孙发现电子的那个时期，他刚好在卡文迪许实验室里学习。

事实证明，卢瑟福是一位不世之材。他在学习期间，对放射性现象的研究令汤姆孙侧目，这部分研究也成为他日后重大发现的契机。实际上，仅仅在汤姆孙1906年因为发现电子而获得诺贝尔物理学奖后的两年，卢瑟福就因对元素的衰变以及放射性方面的研究而获1908年度诺贝尔化学奖。他在1899年及其之后的这些发现也巩固了汤姆孙的观点：原子之中还存在更微观的结构，可以再分，而放射性就是原子衰变出更小微粒的过程，同时还会伴随发射出一些电磁波。

这些发现，也刺激了卢瑟福进一步思考原子的结构问题，他猜想放射线说不定可以用来验证汤姆孙的原子模型。

几乎就在同一时间，日本科学家长冈半太郎（Nagaoka Hantaro，1865年—1950年）提出了一个很离奇的观点。长冈半太郎曾经前往欧洲参加过物理学大会，在听过汤姆孙的报告后，也开始思

考原子的结构问题。因为受到土星环的启发，他就猜想原子有没有可能也有一个核心，而电子在核心外绕着飞，就像土星环上的那些岩石绕着土星旋转？

土星和土星环之间的吸引力是万有引力，如果原子也是这样的模型，那么核心和电子之间的吸引力就应该是电磁作用力。擅长数学的长冈半太郎经过复杂的运算后发现，这种结构居然是可以稳定存在的。于是，他在1905年发表了一篇论文，公开了自己的研究结果，提出一种核模型。他认为，原子是由电子绕带正电荷的粒子组成的。

卢瑟福看到了这篇论文，但他并没有立即回应。

几年后的1911年，卢瑟福终于设计出那个彻底影响人类物质观的重要实验——α粒子散射实验，以证明到底哪一种原子模型是正确的。

在这个实验中，卢瑟福用一个α粒子发射源对着金箔进行照射。α粒子的本质是氦原子核，带有正电荷，而汤姆孙此前研究的阴极射线也被称为β射线，本质上是电子流。这两种粒子是卢瑟福于1899年发现放射性辐射中的两种成分，并加以命名的。此外，还有一种不带电的γ射线，本质上是一种电磁波。这3种射线，都是放射性物质放射出的常见射线，卢瑟福对他们早就了然于心。

如果汤姆孙的观点正确，按照卢瑟福的预测，带有正电荷的α粒子在撞到正电荷的原子实体时，大概就像是子弹打到墙上一样，子弹会贴上去，但也说不定会打下点什么碎片。

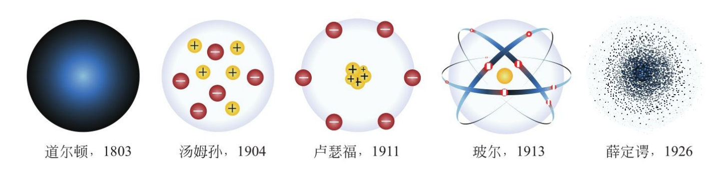
但是，最终结果却让卢瑟福大吃一惊：绝大多数α粒子都如入无人之境一样，直接穿透了金箔，甚至都没有明显的减速。但是，也有少部分粒子的运动方向发生了偏转，还有极少部分的粒子被弹了回去，方向彻底发生了180°大逆转。

这会是什么原因？

尽管不可思议，但卢瑟福还是欣然承认，他的老师错了，长冈半太郎的推测是正确的。只有当原子存在正电荷的原子核、且原子核的尺寸极小时，才会出现实验中的这个结果：金箔很薄，最薄时不过只

有几百个原子厚，而原子内部绝大部分都是空的，所以α粒子什么都不会碰到，直接就撞出去了。不过，也有一些α粒子刚好接近到了原子中心的原子核，因为α粒子和原子核都带有正电荷，两者相互排斥，有些粒子就会因此偏转方向。如果刚好从正面撞向原子核，就会因为强大的排斥力而被弹回来。

就这样，卢瑟福实验发现了原子核的存在，从而推翻了汤姆孙的原子模型，在长冈半太郎假说的基础上，重新构建起一套新的系统，并命名为原子结构的“行星模型”。卢瑟福实验论证原子核的存在，被誉为物理学史上“最美的十大经典实验”之一。对于自己学生的这个做法，汤姆孙丝毫没有感到难堪，甚至在自己从卡文迪许实验室卸任时，还力主由卢瑟福接任。




原子模型的演变

不过，相比于长冈半太郎的土星环模型，汤姆孙模型也并非一无是处。汤姆孙对原子的特性非常熟悉，因此在他的梅子布丁模型中，电子会按照特定的数目进行排列，这样就可以满足“化合价”的需要。对此，卢瑟福不仅进行了继承，还进一步发展了这个模型，简单而形象地勾勒出原子的性质。到如今，出现在中学化学课本上的“原子模型”，其实就是卢瑟福的杰作。这倒不是说卢瑟福的模型就足够完美，我们下一章还将说到，卢瑟福模型中存在一个致命的漏洞，但它毫无疑问是最有助于理解原子结构的一种模型。

不仅如此，这种原子模型也更好地阐释了元素周期律的原理，并且根据周期律，我们可以弄明白原子之间是怎样结合在一起的。

## 原子之间的电磁吸引力

当门捷列夫绘制元素周期表的时候，他只是按照原子量排列出已

知元素——所有的元素被排列成一个矩阵，每一横排都有8个元素（注：门捷列夫原始表格中的横行与竖列与后来的元素周期表相反，并缺少了氦、氖等惰性气体，此处按照现代元素周期表的格式描述），一个横排被称为一个“周期”，一个竖列则被称为一个“族”。

在一个周期内，所有的元素都具有不同的特点，相邻的两个元素会发生渐变。比如从排在11位的钠元素到排在第18位的氩元素，全都排在第三行，是同一个周期，被称为第三周期。第三周期的这些元素，就满足渐变的特征。

钠和镁相邻，它们都是活泼性非常高的金属：金属钠扔在水里，就会发生非常剧烈的反应，产生大量的氢气，而氢气是一种可以燃烧的气体，要是反应不受控制，说不定还会因为温度过高而爆炸；金属镁虽然不会和冷水发生反应，但是放在热水里，它也一样会产生大量的气体。

排在镁后面的是铝和硅，从偏旁就可以看得出来，铝是一种金属，而硅却是非金属。铝也很活泼，可是相比于钠和镁来说，就要差远了。它和水之间的反应很慢，只有在和水蒸气接触的时候才会发生剧烈的反应。至于硅，它就很难和水直接发生反应，又要比铝差了一些。

除了这些反应活性的差异，更直接的渐变在于，这些元素在参与形成分子的时候，化合价也会依次升高。钠的化合价是1，镁是2，铝是3，硅是4……就像是音符一样。事实上，在门捷列夫之前，就有一些科学家提出了元素周期律的雏形，其中有人就是按照音符的规律进行了排列。

的确，当每一周期的元素终结之后，开启新的周期时，又会出现同样的规律。比如第四周期由钾元素开始，它和钠非常相似，也会和水剧烈反应，而且化合价也是1；钙的活性比钾低一些，而化合价就是2，和第三周期的规律非常相似。

如果按照竖列的方向，那么钠和钾排在同一列，属于同族，镁和钙也在同一列，也是同族。不难看出，同族的元素非常相似，它们有着十分接近的化学性质，更重要的是，它们的化合价都相同。

门捷列夫尽管设计出了“周期”和“族”的元素分类方法，可他始终无法对此进行解释，只因为他根本不相信原子还有更微观的结构。而当卢瑟福提出原子模型的时候，门捷列夫已经仙去，他最终也未能等来这个问题的答案。

事实上，卢瑟福的解决思路非常巧妙。在他设计的系统中，原子核就好比是太阳系中的太阳，所有的电子都沿着特定的轨道绕着原子核旋转。靠近原子核的最内轨道，只能容得下两个电子，所以在元素周期表上，第一周期也只有两个元素，分别是氢和氦，它们的电子数量恰好也就是1个和2个。

在第一层排满了之后，电子就开始进入到第二层。在此之后，不管哪一层是最外层，它最多都只能有8个电子，门捷列夫毕生未能破解难题，就隐藏在这个奇妙的数字中。

理论上说，一种元素的原子最外层有多少个电子，那它就是第几族，化合价也是同样的数字。比如钠，它的最外层有1个电子，所以它排在第I族，化合价也是1，由此推算，镁的最外层排了2个电子，那它就是化合价为2的第II族。

也就是说，只要知道了电子的排列方式，很容易就能弄明白化合价的来历。

那么，钠的化合价为1，实际的含义又是什么呢？

原来，当电子在钠原子核的外围旋转时，最外层的电子只有保持在8个时（第一层是两个电子），才会保持稳定。因此，对钠原子来说，它最外层的那个电子就有些尴尬了——它显得有些多余。

于是，钠原子就采取了一个最有效的策略，随时丢掉那个累赘的电子，形成稳定的结构。正如我们此前所说，当钠原子丢掉一个带负电荷的电子后，那它自身就变成了带有一个正电荷的钠离子。由此看来，所谓的化合价为1，实则就是钠离子的一个正电荷。

这样的规律也的确体现在镁和铝上。镁的最外层有两个电子，脱落两个电子的难度显然比脱落一个更难，所以要想让它变成带有两个正电荷的镁离子，自然也更不容易。于是，镁的活性就要比钠要弱一

些，以次类推，铝又比镁更弱一些。

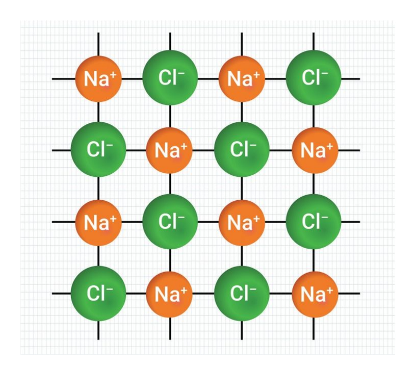
不过，当这个规律继续延伸，一直推到排在第17位的氯元素时，情况又有了新的不同。

氯的最外层有7个电子，它在结合成分子的时候，最高的化合价也的确就是7，但是在大多数时候，它的化合价却只有1而已。原来，氯和钠采取了一个完全相反的策略。对它而言，想要把7个电子全部脱落自然是极其困难，但是，如果在7个电子的基础上再多凑1个电子，很容易就能满足8个电子的稳定结构了。因此，通常情况下，氯都会再多获得一个电子，变成带有一个负电荷的氯离子，化合价也就是1了。

这样一来，原子之间到底怎么结合的问题也就很好解释了。

比如钠和氯相遇，它们会发生化学反应。钠原子倾向于脱落一个电子，而氯离子倾向于得到一个电子，双方各取所需，于是形成了正电荷的钠离子和负电荷的氯离子。这时候，异性相吸的电磁作用力也开始发挥效力，两个带有相反电荷的微粒紧紧地靠在一起，形成了氯化钠（NaCl）。氯化钠也就是生活中常见的食盐的主要成分。

在实际过程中，参与反应的原子不会只有一两个，而是数以万亿计。当无数个钠离子和无数个氯离子相遇时，它们就会彼此交错地堆积起来，每一个钠离子的周围都是氯离子，同样，每一个氯离子的周围也都是钠离子。虽然每一个离子能够产生的作用力都很微小，但是因为这样的离子实在是太多了，它们产生的结合力就非常惊人。如果我们从厨房找出几粒粗盐，想要把它们研碎——千万别尝试“摧心掌”之类的武学秘籍去和它硬碰硬，它锐利的棱角足以把手上的皮肤割破。




氯化钠模型

正因为这种结合力是靠着电荷的电磁吸引力实现，所以电荷越大的离子，就可以实现特别强大的性能。比如地球上普遍存在的氧化铝（$Al_2O_3$），主要成分就是正电荷的铝离子和负电荷的氧离子，其中铝的化合价是3，而氧的化合价是2，它们之间的吸引力远比氯化钠更强。结果，想要把它们分开可就太难了。在地球上，天然形成的氧化铝，硬度极高，是红宝石和蓝宝石中的主要成分。如果想要把这种物质从固体熔化成液体，就需要升到很高的温度才可以，否则不足以打破铝离子和氧离子之间的吸引力，氧化铝也就无法流动起来。

可见，不同的原子可以靠着电磁作用力相互吸引，结合成更大的结构。然而，回到我们第3章的那个问题，相同的原子又该如何结合呢？回答这个问题，我们将会领略更普遍存在的物质作用力。

## 同性因何不相斥？

异性相吸而同性相斥——这是最朴素的电磁学理论告诉我们的现象。

然而，当科学不断发展之际，却有很多新的现象让人感到奇怪。比如，我们都知道，原子核与电子分别带有正电荷与负电荷，它们因为电性相反而吸引，于是电子绕着原子核转。与此同时，我们还知道，原子核是由正电荷的质子与不带电的中子结合而成，在那么细小的结构中，都是正电荷的质子又是怎样结合在一起的呢？

现代物理学证明，在非常近的距离下，质子之间还有一种被称为“强相互作用”的力，是它将多个质子锁定在一起。这种力我们在日常生活中绝对感受不到，因为它只在原子核那么大的空间里起作用。甚至当原子核变得更大一些，比82号元素的铅原子核更大时，强相互作用力就因为作用距离太远而消失殆尽。于是质子与质子之间相互排斥的电磁作用力占据主导，原子核就倾向于裂开成小一些的原子核——这就是核裂变的过程。反过来，当氢原子核靠得足够近时，强相互作用力会让这些原子核发生融合，核聚变就发生了。

对原子核的研究还发现，在极短的距离下，还有一种弱相互作用力，它和强相互作用力一道，左右着原子核乃至更小微粒之间发生的很多行为。这两种作用力，和我们早已熟悉的万有引力以及电磁力一道，被称为物质世界的四大基本作用力，物质就是靠着它们组织在一起。

不过，物理学家对于这个观点仍然不满意，很多人坚持认为，四大基本作用力应当被统一起来，它们有着一样的本质。经过不懈的努力，弱相互作用力和电磁作用力已经得到了统一，而强相互作用力也取得了一定的研究进展，只有万有引力显得特立独行。

如果把眼光拉回到我们的生活中，不难发现，强相互作用力与弱相互作用力的距离实在太短，我们用不到；而引力虽然无处不在，但它的作用系数实在太小，能够让我们产生切身体会的，也就是地球自身的引力——若是观赏钱塘江大潮，倒是可以亲眼看到月球引力的影响。

这样一来，电磁作用力就成了我们生活中最普遍存在的一种作用力，某种程度上说，也可以说是影响面最广泛的一种力，因为原子之间的各种作用力也都以此为基础，是它让我们身边的物质世界发生着各种变化。

然而，有很多气体单质，比如氢气、氧气、氮气，它们都是由相同的原子结合在一起——它们本应该具有相同的电性，为何没有同性相斥呢？正如我们在上一章所提到的，“分子”学说之所以被冷落了半个世纪，主要就是因为这一点矛盾。

19世纪初，就在原子论刚刚被提出来的时候，虽然科学家对原子结构的研究还为时尚早，但是当时一些科学家还是凭经验领悟到了原子结合过程的内涵。这其中，最有影响力的莫过于瑞典科学家永斯·雅各布·贝采尼乌斯（Jons Jakob Berzelius，1779年—1848年）提出的电化二元论，这种理论敏锐地指出，不同的原子能够结合，就是因为能够形成不同的电性，就像上面讲到的氯化钠那样。在不知道原子结构的前提下就能够做出这样的论断，的确很了不起，它也有效地解释了很多物质为什么会存在。

然而，随着分子学说被重新提起，特别是氢气这类由相同原子结合起来的分子被证实是真实存在的，电化二元论开始走向破产，可新的理论却又迟迟没有建立——这些简单分子引起的学术争论，直到20世纪中期才被平息。

电化二元论说对了一半。

像氯化钠这样的一些物质，从现在的观点来看，已经不能被称作“分子”。正如前面所说，它们是由原子首先变成离子，再由无数个正离子与负离子交错搭建，不是阿伏伽德罗猜测的“小团体”模式。实际上，尽管我们称之为氯化钠，用NaCl这样的化学式去指代它，但是在食盐中我们根本找不到只由一个钠和一个氯结合起来的“分子”。如今，我们通常用“离子化合物”来称呼这类物质，一眼就能知道它们的特征。

而当电化二元论被用来解释更多物质的时候——比如水——就夸大了电荷吸引力的程度。

水分子是由两个氢原子与一个氧原子构成，其中，氢的最外层只有1个电子，而氧的最外层有6个电子。可以猜到的是，氢更容易失去电子形成正离子，而氧容易得到电子变成负离子，于是它们异性相吸，正负结合。

但是，氯离子和钠离子之间的作用力，可以让氯化钠直到800摄氏度时才会熔化。水在常温下却是液态，尽管它的存在还有固态（冰）和汽态（水蒸气）的聚集状态，但无论怎么看，水都不像是由离子聚集而成。

随着原子结构越来越清晰，“化学键”的概念被提了出来。它是分子或原子团中两个或多个原子（离子）之间因强烈的相互吸引而结合在一起的作用。对于水分子这样的结构，也有了更可靠的解释。价键理论就是关于化学键的基本理论，它也可以用来解释元素的化合价。这是海特勒（Walter Heitler，1904年—1981年）和菲列兹·伦敦（Fritz London，1900年—1954年）于1927年用量子力学处理氢分子所得结果的推广和发展。

原子与原子之间的作用力，就像是无形的锁链将原子扣在一起，而“键”的本意就是“锁”，因此，把这种作用力称作“化学键”实在是太契合不过了。

像氯化钠这样的物质，它的化学键实际上就是正负离子之间的静电引力产生的，所以被称为离子键。在NaCl这样的盐类晶体中，可以很容易找到离子键。除了离子键，还有一种被称为“共价键”的化学键，它是电磁作用力的另一种表现形式。共价键就是两个原子结合时，通过共享电子对而形成的化学键。

共价键的形成过程，是两个原子相互接近时，不需要发生极端的电子迁移，而是以共享电子的形式结合在一起。这样一来，所有参与的原子都可以形成8个电子（或两个电子）的稳定结构。就说氢气吧，两个氢原子各有一个电子，它们结合在一起就有了两个电子，它们同时绕着两个氢原子旋转，于是每个氢原子都有了稳定的结构。

打个比方，离子键就像是借贷关系，钠的电子借给了氯，它们从此绑定在了一起；共价键则好比是夫妻关系，各自拿出一部分电子作为共同财产过日子，如胶似漆。

像氧这样的原子可以形成两个化学键，所以当氧和氢碰到一起时，就需要两个氢原子才能和一个氧原子结合，这才有了水分子。

通过化学键的形式，形形色色的原子都可以连接在一起，而它的本质却仍然是以电子为载体的电磁作用力，和我们在宏观层面上看到的玻璃棒吸引纸片有着莫大的相似性。1938年，莱纳斯·鲍林（Linus Carl Pauling，1901年—1994年）出版了《化学键的本质》，完整地阐述了这些观点。这部作品引起了巨大的轰动，终结了物质在原子层

面如何连接的世纪之争，也成为鲍林1954年荣获诺贝尔化学奖的重要依据。

也正是对化学键有了足够的了解，我们可以操控原子，让它们按照理想的结构进行排列。然而，我们怎样才能对原子施加这些力？

人不吃饭就没有力气，这个道理对所有的物质都成立，因为我们需要耗费能量才能进行各种操作，而物质和能量本就是一体的。下一章，我们就来说说它们之间的关系。

# 5 永不消失——物质和能量是怎样转化的

## 奇妙的等式

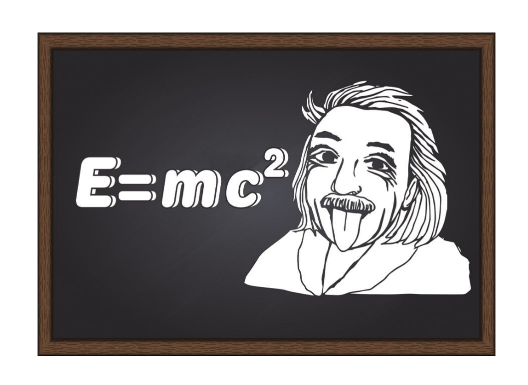
1945年8月6日和9日，两枚原子弹分别落在了日本的广岛和长崎，数十万人当场丧命，在后来的几十年里，因为核辐射而遭受身体与心理双重伤害的人更是不计其数。几天之后，日本天皇宣布无条件投降，第二次世界大战至此落幕。

在人类的历史上，战争与和平一直是最重要的议题之一，战争烈度的不断升级，也在刺激着科学技术的飞速发展，原子弹更是用血淋淋的事实证明了这一点。

多年以后，随着“曼哈顿工程”的各种细节不断解密，人们得以串联起20世纪上半叶的很多历史瞬间，从而逐渐明白，除了在战争方面的巨大影响外，原子弹成功被引爆，也刷新了人类的物质观。科学本无善恶之分，如何利用科学为人类谋幸福才是科学的真谛。

我们已经知道，在太阳内部，核聚变已经持续了至少45亿年。所谓核聚变，就是轻原子核聚合变成较重的原子核，同时释放出巨大能量的过程。原子核剧烈撞击并融合的过程，以氢和氦的聚变为起点，新的元素因此源源不断地生成，我们此前已对此过程有所了解。原子弹所用的原理与之相反，被称为核裂变，就是原子核分裂为两个质量相近的核（裂块），同时释放出中子的过程。较大的原子核在此过程中会分裂成更小的原子核，在此过程中也会生成新的元素。

核裂变可以用于制造原子弹，而在原子弹首次得到应用的九年后，以核聚变为原理的氢弹也试爆成功了。这两种核弹，背后都离不开一条重要的公式：$E=mc^2$。公式中的$E$代表能量，$m$是物质的质量，$c$代表光速。这条美丽的公式最初由爱因斯坦于1905年提出，通常被称为“质能方程”。




理论上说，当一种物质消失的时候，它的所有质量都会转化为能量，并且总能量可以由“质能方程”计算得到。只看这个公式，它似乎是在说：如果我们找到一块木头，又挖出一块煤球，它们刚好都是1千克。尽管木头和煤炭的成分不同，所含的各种原子数量也不同，属于不同的物质。然而，它们具备相同的质量，也就蕴含着相同的能量。

直观来看，这是有违常识的。就像木头和煤球，如果用来给炉子加热，煤球可比木头耐烧多了，显然，煤炭富含更多的能量。

然而，当我们以更普遍的角度来看待“能量”的时候，就会理解爱因斯坦的世界观，也能更深刻地知晓物质的规律。

200多年前的18世纪，以法国为中心，上演了一场有关“燃素”的辩论。这场规模盛大的辩论，几乎吸引了当时所有最出名的自然科学家，特别是诸如法国的安托万-洛朗·拉瓦锡（Antoine-Laurent de Lavoisier，1743年—1794年）、瑞典的卡尔·威尔海姆·舍勒（Carl Wilhelm Scheele，1742年—1786年）、英国的约瑟夫·普利斯特里（Joseph Priestley，1773年—1804年）等一些顶尖化学家。所谓“燃素”，当时认为，它是一种支撑物质燃烧的过程中存在的“元素”，燃烧时燃素以光和热的形式逸出，物质的质量在燃烧之后一般也减少了，因此燃素好像也有质量。当一种物质含有燃素时，它就可以燃烧，而当燃素被脱除后，它便不再能够燃烧。

以后世的眼光看，如此众多的科学家汇聚一堂，只为研究燃烧的过程，这多少有些小题大做。在物质科学史上，很多故事都是这样，20世纪的量子科学也只是从“黑体辐射”这个现象开始的。

尽管后来遭到批判，然而诞生之初的“燃素说”，很大程度上却是合理的，因为它符合我们观察物质的第一视角。就像前面说到的木头和煤球，要想理解同样质量的煤球为何能够在燃烧时释放出比木头更多的热量，那么只要做一个思想实验，想象其中蕴含某种可以燃烧的元素，煤炭所含的燃素比木头更多，那么这个问题就迎刃而解。

不仅符合观察结果，燃素说甚至还有其进步之处。世界上可以燃烧的物质很多，它们的形态各异，燃烧时的状态也不尽相同，而燃素理论将燃烧的现象归纳为更普遍存在的化学反应，有助于解释整个物质世界的一般规律。

只是在解释更多的现象时，这一理论出现了矛盾。

举个例子，燃素说的支持者注意到，木头和煤炭燃烧之后会变轻，这似乎可以理解成燃素逸出后的结果；然而，有一些金属也会燃烧，并且它们燃烧后的质量是增加的，这实在令人匪夷所思。

对于这个现象，燃素说的理论家们提出了一个足以让燃素说寿终正寝的荒诞推论：燃素可以是负质量的物质。也就是说，有些物质在燃烧之后，其质量不降反增，是因为它们在这种情况下释放出了负质量的燃素。

这种自相矛盾的说法，令包括拉瓦锡在内的很多科学家下决心进一步探索。拉瓦锡最终证明，物质燃烧和动物呼吸的本质是氧化反应，据此驳斥了不正确的“燃素说”。有的物质燃烧后的质量增加，是因为氧气或其他氧化剂成了燃烧灰烬中的一部分；有的物质燃烧后的质量减少了，只是因为燃烧中产生的二氧化碳、二氧化硫等气体没有被收集，实质上它们的质量也是增加的。至此，燃烧氧化说终于取代了错误的燃素说。

在此基础上，拉瓦锡还乘胜追击，提出了“质量守恒定律”。如今，化学教科书会在最开始就讲述他的这段故事——的确，质量守恒定律是支撑化学这门学科最核心的基础之一。

问题到这里并没有结束。事实上，质量守恒定律并没有解释所有问题。还是以木头与煤球为例，如果木头和煤球在燃烧后，再加上参与反应的氧气，质量也都没有发生改变，为什么它们释放的能量却不

一样多？

或者我们还可以思考几个更简单的问题：一个铁球，放在山顶上所蕴含的能量和放在山脚下时一样大吗？显然，我们都知道，山顶上的势能更大，铁球蕴含更高的能量，那么在其他参数都不变的前提下，山顶上的铁球质量和在山脚下一样吗？进一步思考，被压缩的弹簧，飞出枪口的子弹，它们都各自累积了势能或动能，是否和此前一模一样呢？

这些略显得荒唐的实际问题，或许想破脑袋也不知道从何处着手。不过，随着爱因斯坦提出相对论之后，物质世界就发生了剧变，而相对论也是理解这些奇妙问题的绝佳工具。

我们已经知道，光是一种电磁波，麦克斯韦方程描述了它的运动过程。很奇怪的是，如果根据麦克斯韦方程进行计算，光速是不变的。这里的“不变”，说的是不管观察者自己的状态如何，看到的光速都是一样的。好比说，一道闪电出现，乘客不管是待在站台还是坐在高速行驶的列车上，他们看到闪电发出的光，速度是一样的。

如果不是闪电的光而是雨滴——当站台上人看到雨滴垂直落下时，火车上的人将会看到雨滴向斜后方运动，二人看到的雨滴速度并不相同。看起来，光速不变的特性与我们的常识并不吻合。

然而，尽管这个计算结果这么离奇，它还是被实验证明了。1887年，两位科学家完成了一次著名的实验，以至于后人冠上他们的名字，称之为迈克耳逊-莫雷（Michelson-Morley）实验。正是这个实验，证明光速不变是真实存在的规律。

在牛顿力学范围内，时间和空间的测量与参考系的选取无关，这就是时间的绝对性和空间的绝对性。有了光速不变的实验基础，爱因斯坦重新思考了那个火车与闪电的思想实验。我们都知道，在匀速直线运动中，运动距离$s$等于运动时间$t$和速率$v$的乘积，即$s=vt$。当闪电的光传递到移动中的火车上和火车旁静止的人时，假如闪电发生的那一刻，闪电与两个人是等距的，但是因为相对运动的关系，两个人看到闪电的时候，各自与闪电击中的距离却是不一样的。又因为光速相对于两个人的速度不变，也就是说，在上面那个公式中，$s$变了，$v$

没有变，那么只有一种结果，那就是$t$变了——两个人的时间不同。更直接地说，运动的那个人，时间相对流逝得更慢，这就是狭义相对论的基本理论。物理规律对所有惯性参考系都是一样的，不存在任何一个特殊的（如“绝对静止”的）惯性系，即物理定律对所有惯性系都是等价的。针对牛顿绝对时空观存在的问题，爱因斯坦建立了物理学中新的时空观和（可与光速比拟的）高速物体的运动规律——狭义相对论。由于涉及的只是无加速运动的惯性参考系，所以称为狭义相对论，以区别于后来爱因斯坦推广到非惯性参考系的广义相对论，他在那里讨论了加速运动的参考系。

在这个基础之上，爱因斯坦又进一步思考，得出引力对于时间的影响，也就是在引力很强的位置，时间也会变慢，这又是广义相对论的基本理论。

根据相对论，可以得出很多推论，其中一条就是质量的变化——运动的物体质量会增加。不仅如此，任何让物体能量增加的行为，都会体现在质量上。也就是说，理论上讲，压缩的弹簧会比松弛的弹簧质量更大。至于质量会变化多少，爱因斯坦通过缜密的计算，得出$E=mc^2$这个公式。

通过这个公式，我们可以得知，当物体的质量消失时，它可以转化为巨大的能量。但是，因为这个公式中的光速（$c$）达到了大约每秒30万千米，以至于哪怕只有非常细微的质量变化，都会造成极大的能量变化。而在生活中，就算是炸药爆炸这个过程，产生的能量已经很大了，可质量变化还是小到根本不能用仪器测量出来。

在爱因斯坦研究相对论的那段时期，人们所知的过程，只有一种有可能出现明显的质量变化——包括聚变与裂变在内的核反应。这些核反应都会引起质量减少，而它们释放的能量，比其他任何形式都要大得多。如果把这些反应放到军事武器中，那可就了不得了。

于是，人类史上最大胆的科研攻关计划在美国出现了，那就是以核反应为基础制造出原子弹的曼哈顿工程。论证了质能方程的爱因斯坦也应邀参加了这项工程，然而当他知道这项工程的目的后，信奉和平主义的他并没有继续下去。

他很清楚，一旦原子弹的研究取得成功，将会造成多大的杀伤力。最后在日本上空爆炸的两枚原子弹名叫“小男孩”和“胖子”，它们所用的爆炸材料分别是“铀”和“钚”。以“小男孩”为例，它所含有的铀只有区区64千克，和一名成年人体重相当，爆炸后减少的质量更是不足1克，但是却能破坏整座城市。

这样看来，前面说到的1千克木头或煤球，如果真的能够彻底转化为能量，那么按照质能方程算下来，会是一个令人恐惧的结果，人类目前也还没有技术能够做到这一点。而它们在燃烧时也的确会出现质量变化，只是微乎其微，所以“质量守恒定律”通常还是适用的。

## 被激发的电子

在物理学中，1905年被称为“爱因斯坦奇迹年”，甚至100年后的2005年，还专门为此设立了“国际物理年”。因为爱因斯坦在这一年发表的5篇论文，每一篇都可以说是开创性的伟大工作。1905年，爱因斯坦提出的“相对论”可以说是石破天惊，10年后他又进一步完善，提出了“广义相对论”，构建出一个完全超乎想象的物质世界。然而，这对当时的人们来说，实在是过于超前了，以至于科学界最权威的人士都不敢贸然判断他的说法是对还是错。事实证明，这样的犹豫是有必要的，就像阿伏伽德罗提出分子理论时，正是因为主流科学界因为不理解而弃之如敝屣，以至于物质科学的发展被大大地延后了。

当时，爱因斯坦的思想远远超前于那个时代的所有科学家，除在数学上曾得到马塞尔·格罗斯曼（Marcel　Grossmann，1878年—1936年）和戴维·希尔伯特（David　Hilbert，1862年—1943年）的有限帮助之外，几乎单枪匹马奋斗了9年。爱因斯坦曾自豪地说：“如果我不发现狭义相对论，5年内就会有人发现它。如果我不发现广义相对论，50年内也不会有人发现它。”到了1921年的时候，诺贝尔奖评奖委员会坐不住了，爱因斯坦对科学的贡献实在是太大了，如果一直不颁奖给他，怎么也说不过去。

然而，还是有些物理学家不同意这样的做法，巨大的争议之下，1921年的诺贝尔物理学奖出现史无前例的轮空。直到一年后，另一位

大科学家尼尔斯·玻尔（Niels Bohr，1885年—1962年）获奖，而玻尔获奖的理由却与爱因斯坦的其中一项“奇迹”相关，于是诺贝尔奖委员会顺水推舟，以此为由补发了爱因斯坦的诺贝尔奖，填补的是1921年度诺贝尔物理学奖的空缺，但是他获奖的理由却不是创立相对论。爱因斯坦拿到获奖的电报时，刚刚从香港到上海的轮船上登岸。

爱因斯坦获奖的这项工作，就是鼎鼎大名的“光电效应”，它很好地解释了能量，特别是光这种能量与物质之间的联系。

简单来说，光电效应指的是光照促使物质形成电流的过程。这种现象如今已经被广泛地应用在太阳能光电板中：在太空站上，在戈壁滩上，在农村的大屋顶上，到处都可以看到这种装置。它可以稳定地吸收太阳光中的能量，让物质中的电子运动起来，形成电流后传递到电池中，光能于是以电能的形式被储存起来，用于驱动各种电器。

早在1887年，德国科学家海因里希·赫兹（Heinrich Hertz，1857年—1894年）就已经在实验过程中观察到光电效应，但是对于这种现象发生的过程，却没有办法解释。正如第4章所说，光是一种电磁波，没有质量，电子却是一种有质量的物质，那么电磁波又怎么能推动电子运动呢？事实上，在光电效应中，还有很多奇特的现象，科学家们修正了很多物理学理论，依然不能将其中的道理讲清楚。

爱因斯坦首先设想，光或许还有“粒子”的特性，像一个个小球，这些小球在静止的时候没有质量（静止质量为零）。正如前面质能方程所计算的结果，运动中的光既然具有能量，那么它就增加了质量，如此一来，它不就可以对电子产生作用了吗？他把这种粒子称作“光子”，并且强调这种光子的能量与光波的频率相关，频率越高（波长越短）的光能量越高。这样一来，这些光子就像切菜一样，把光的能量分成了一份又一份，所以光子也常被称作“光量子”，这种思维体系则被称作量子力学。

这样一来，光电效应中一些原本看起来很奇怪的现象就能够被解释了。比如，用光照射某种材料的时候，如果紫色的光可以产生电流，红色的光就不能产生电流，那么，无论紫色的光多弱，红色的光多强，都不会改变这个结果。

如果认为光波中的能量是由一个个光子承载的，紫色的光频率高而波长短，每一个光子的能量就大，红色光与之相反，每一个光子的能量就小。电子就像是落在坑里的一只足球，要想让它滚动起来，就需要有足够的能量把它从“坑”里踢出来。紫色的光子能够做到，红色的光子却做不到。紫色光弱，光子的数量少，但也只是踢出的电子少了一些，红色光再强，光子再多，它也仍然不能让电子形成电流。



现代光伏电站也利用了“光电效应”

爱因斯坦提出的这个思想，也启发了玻尔对原子结构的思考。

第4章提到，卢瑟福的原子结构简单而直观，但是在描述原子的时候，也存在一个非常致命的漏洞——如果负电荷的电子绕着正电的原子核旋转，那么它为什么不会释放出能量，直到电子和质子吸引在一起了呢？也就是说，卢瑟福行星模型无法解释原子的稳定性和原子有一定的大小。诚然，卢瑟福行星模型存在一定的局限性，但由于其直观且容易理解，所以在不少情况下，仍用以作为对原子结构的一种粗浅说明。下面讲的能级的概念等纯粒子性的表述，也是因较为形象和易于计算，至今都在沿用着。

玻尔猜测，电子在原子核外的不同轨道上运动，每一个轨道都有各自的“能级”。所谓能级，就像是能量构筑的台阶——相邻的台阶之间，总是有着特定的能量。电子在原子核外运动时，就处于不同的“能级”。如果电子都处于最低的能级时，物质这时的状态就被称为基态。而当电子接收到能量后，就会发生跃迁，像是爬楼梯一样，爬

到“激发态”。所有电子都只在基态上的物质，实际上只存在于理想当中，因为我们找不到完全不携带能量的物质。

这样一来，原子核外的这些电子，吸收的能量也是“量子化”的，它必须要吸收或者释放特定的能量才会发生“跃迁”。我们知道，地球以外绕转的那些人造卫星，报废以后就成了太空垃圾，它们会慢慢地减速，逐渐消耗自身的能量，最后落在地球上。但是，绕转原子核的电子就不会这样，因为它不能“逐渐”消耗能量，只是一直在轨道上旋转。

而当电子要从“基态”跃迁到“激发态”的时候，就像爬楼梯一样，它也必须每一次都爬升特定的“能级”。光电效应是一种特殊的跃迁，它的能量足够高，让电子直接爬到了楼顶，变得“自由”了——就像坑里的足球被踢出来了一样。

玻尔的这些创造性思想，如今被称为原子的“玻尔模型”，它也在很多实验中得到了证明——比如光谱实验，我们下面还会讲到。不过，这种把电子轨道“量子化”的处理方法，也让物质之间的化学反应得到了解释。

为什么有些反应必须要达到了一定的温度才会进行呢？因为加热是输入能量，此时承载能量的微粒也是光子，只是它们处于红外波段，比可见光的频率更低，我们肉眼看不到。但是不管怎么样，这些光子撞到了电子，让电子发生了跃迁，进入了激发态，运行的轨道比原来更远离原子核。当两个这样的原子撞到一起的时候，它们就更有机会发生结合，当然，如果已经结合在一起的原子，也会因为这个过程更容易分开。于是，它们在这种状态下找到各自更契合的原子伴侣，化学反应也就完成了。

实际上，玻尔的这个模型，对爱因斯坦的相对论也是很好的补充。根据计算，因为有些轨道的能级实在太高，电子在这些轨道上绕转的时候，运动的速度实在是太快了，以至于有些电子的质量比静止状态时高出了20%以上。这样一来，我们不只是可以窥探核反应过程中质量转化为能量的原因，还可以计算出来，这些电子与原子核之间的吸引力发生了很大的变化。这种变化效应被称为“相对论效应”，它

可以被用来解释很多，比如为什么钨的熔点那么高（约3 400摄氏度，所有金属中最高），汞的熔点又那么低（-38.87摄氏度，所有金属中最低），而锇的密度又会那么大（22.59克每立方厘米，所有元素中最大）。

必须指出，玻尔理论对只有一个电子的氢原子和类氢原子的谱线频率作出解释无疑是成功的；海森伯的位置与动量不确定关系表明，玻尔模型不能正确地描述电子在原子中（如多电子原子）的行为，也不能说明谱线的强度和偏振等现象。玻尔假设（玻尔模型）属于半经典半量子的理论，尽管后来经德国物理学家阿诺德·索末菲（Arnold Sommerfeld，1868年—1951年）等人的修改和推广，但仍未能摆脱困境。尽管如此，玻尔理论的部分成就，促进量子论的发展，在科学史上曾起很大作用。在探索真理的过程中，理论上的缺点是难以避免的。随着科学探索不断深入，我们期待有更优的模型超越玻尔的理论。

在20世纪的物理学历史上，爱因斯坦和玻尔之间曾发生过一场旷日持久的论战。不过，他们并不是“敌人”，而是“战友”，正是他们对相对论和量子力学的探讨，让我们现在对于物质世界有了更清晰的认识。

接下来，我们就来看看，人类是怎么“看”物质的。

## 看清物质

物质世界精彩纷呈，我们睁开双眼就能看到各式各样的物质。在公园里，我们可以远望蓝天白云，也可以欣赏湖面波光，又或是端详野草怪石，偶遇几个亲戚朋友。所有这一切都是由物质构成的，这一点我们早就知晓，但我们是怎么辨识出不同物质的呢？

最重要的辨别方式，自然就是看——不同的物质可以发射或反射出不同的光，我们的双眼正是通过对这些光进行反应，才最终识别出它们是什么，或者远远就认出熟悉的人。

这种作用方式的基础，就是物质和能量的相互关系，而我们今天

能够借助的各项仪器，绝大多数也是利用了类似的原理。

这其中，最常用的一种方法被称为“光谱分析”，此前我们说到从太阳中看到氦元素，其中利用的方法就是通过对其辐射的特征谱线进行分析鉴定得出的。

光谱，顾名思义就是依光的波长大小排列的谱图，它是记录某种物质发射或吸收光波的一种图案，根据光谱去识别物质，就跟看着乐谱唱出歌曲一样。

光谱的形式非常多样，因为物质发生能量变化的方式实在太多了，电子跃迁只是最常见的一种，我们双眼通常也是根据这一点发挥作用。

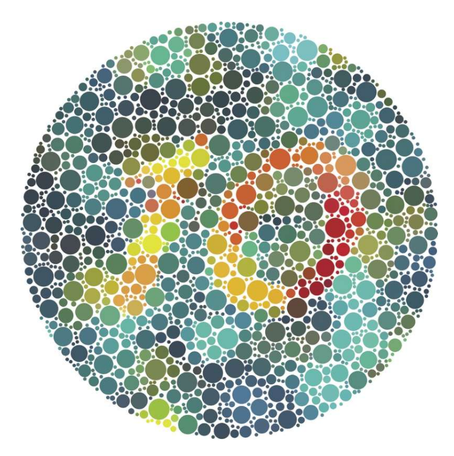
比如，当我们去欣赏月季花的时候，五颜六色的花朵屹立在绿叶环抱的枝头，而我们的眼睛可以真切地看到每一朵花，这就是一种可见光形成的光谱。在这个过程中，观测设备是我们的双眼，而识别设备则是大脑中的处理系统。

在这些月季花和它们的叶子中，含有很多不同的色素分子。这些分子是由原子构成的，每一个原子中的电子，在吸收光波后，就会发生跃迁。正如前面已经提及的，电子的跃迁必须要遵循一定的能级，是量子化的。所以，当太阳光照射过来的时候，那些能量没能满足电子刚好在能级之间发生跃迁的光子，就不会被吸收，通常会直接反射出来。这样一来，我们所看到的颜色，就是色素分子在吸收过所需光波后剩下的那些颜色。

眼睛本身也是一个能量与物质相互作用的场所。就人眼而言，大多数拥有的视觉系统中都有三种被称为“受体”的感光结构，它们分别带有一些物质可以和不同颜色的光发生作用，然后再把信号传递给大脑，大脑由此识别出红、绿、蓝三种颜色。实际上，人类现在绝大多数电子设备也都是这样设计的，比如电脑屏幕，也可以发出红绿蓝三种颜色，它们可以按照不同比例，调和成包括白色在内的各种可见光。

如果人类只能看到两种颜色，那么我们可以识别的物质将会大为减少——事实上也的确有很多人天生色盲，缺少某种感光成分，对生

活产生了诸多不利（“色盲”这种现象最早也是由道尔顿发现的，因此色盲症又有“道尔顿症”的称呼，道尔顿的实验水平相对不高，与此可能有一定关系）。



看一看，图中的数字是什么？

相反，如果感光的成分更多，就会更容易观察到物质的变化。鸟类和一些爬行动物的眼睛中都有着四种或更多的受体，它们看到的世界也比我们人类更丰富多彩。

不过，在掌握了能量与物质相互作用的规律后，人类用检测仪器弥补了这些不足。

很多物质的电子跃迁并不在可见光区，也就是说，它们吸收的光是我们所看不到的，于是这些物质在我们看来，要么就是平平无奇的白色、灰色或黑色，要么干脆就是透明的。玻璃就是一个很好的例子，我们偶尔会莫名其妙地撞到玻璃上，就是因为它看上去空无一物。然而，这并不代表玻璃的内部就不存在电子跃迁，只是因为它们吸收的光波非常短，是比可见光频率更高的紫外光，对可见光却没有兴趣，所以我们看到的就是像水一样晶莹清澈的玻璃。如果我们肉眼能够看到紫外光，那我们看到的玻璃或许就和青铜器差不多了。

对于这些肉眼根本无法分辨的物质，紫外光谱仪器就可以轻松地看出它们的差别，它就是我们眼睛的延伸。

除了紫外光谱以外，还有红外光谱、拉曼光谱、荧光光谱之类的各种光谱仪，都可以帮助我们看清物质。它们全都是利用了电子在能量作用下发生跃迁的原理。即便只是在可见光区，有很多物质也需要仪器的帮助才能看明白它们的真身——这也正是我们从太阳光谱中找到氦元素的办法。

再说一下光。从本质上讲，我们所说的光指的是可见光，从紫光到红光区域。广义上，光不仅是可见光，还包含红外线和紫外线等。虽然红外线、紫外线以及在它们波长之外的电磁波均不能引起人眼视觉，但紫外和红外波段的电磁波可有效地转换为可见光，利用光学仪器或摄影与摄像的方法可以量度或探测发射这种光线物体的存在，因此，在光学研究领域，光的概念通常延伸到邻近可见光区域的电磁辐射（红外线和紫外线），甚至X射线等也被认为是光。

但是，能量对于物质世界的影响还远不止于此，我们的生活，无时无刻不在感受着能量的变化。

## 物质的状态

我们已经知道，水有固态、液态、气态这三种状态，在温度条件不同的时候可以发生改变，而液态的水对于地球生命而言非常重要。

物质的状态和能量有着直接的关系，但它同样也会受到其他很多因素的影响，比如压力。在地球海平面，水在100摄氏度时会沸腾，液态的水蒸发成水蒸气；但是到了高山上以后，因为气压变低了，水的沸点也会下降，可能80摄氏度就沸腾了；要是反过来，用高压锅煮开水的话，水的沸点就会增加，一直到120摄氏度左右才会沸腾。

如果压力再大一些的话，液态的水会更难沸腾，甚至直到很高的温度，水也不会彻底变成气体，而是形成一种不像水也不像气的物质——这种状态被称为超临界流体，它具有很多特殊的性质。

有一些物质比较容易出现超临界流体，二氧化碳就是如此。在地球表面自然的环境下，二氧化碳并不存在液态——如果给这种气体降温，会在零下78摄氏度的时候直接转化为固态二氧化碳，也就是我们

前面讲到的“干冰”。不过，只要在常温条件下给它施加很高的压力，大约是高压锅内最高压力的10倍左右，就能获得二氧化碳的超临界流体。这种流体的溶解能力非常强，可以被用来提纯很多物质，如今已经应用在医药、食品等领域广泛应用。

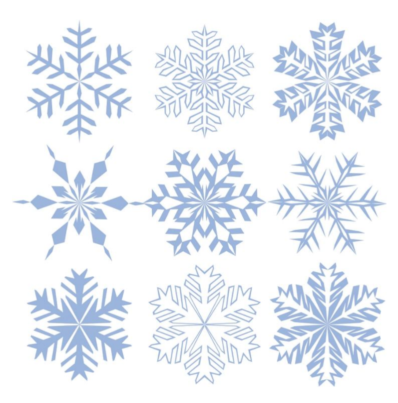
为什么流动态的物质——液态的水或超临界流体二氧化碳——会有如此特别的作用？这还是能量与物质交换的规律决定的。

当外界温度非常低的时候，所有的物质都有可能凝固。氦是所有物质中熔点最低的，只有不到1开，也就是低于-272摄氏度。当物质处于固态时，所有的原子就像排队做操一样，停留在各自确定的位置上，尽管也会小幅度振动，但是相对位置却不会改变。

随着温度上升，这些原子的振动幅度会持续增加。当然，正如前面所说，能量也会被电子吸收，让电子跃迁到更高的能级。如果此时的电子跃迁就引发了化学反应，那么这种物质就不会转化为其他状态，而是直接在固态时就分解了。有一种优质的化肥叫碳铵，它的实际成分叫碳酸氢铵（$NH_4HCO_3$），只要在阳光下照射一会儿，它就会发生分解，变成氨气、二氧化碳和水，最后什么固体都不会剩下。所以，这种肥料见效很快，失效也很快，用起来还需要点技巧。

不过，大多数物质还是可以支撑到熔化的时候，熔化时的温度就被称为熔点。并不是所有的物质都有熔点，玻璃就是这样。有固定熔点的物质被称为晶体，比如冰、铁、食盐、石英都是这样，而像玻璃、松脂、沥青之类的物质就被称为非晶体。

如果从原子层面上看，晶体和非晶体的区别就更有意思了。



常见的雪花形状

晶体中的原子排列非常整齐，它们形成了“晶格”，也就是由原子在晶体内部形成的格子。它们之所以能够排列成格子，也和玻尔设想的原子结构模型有很大关系。当原子外的电子必须按照特定的轨道运动时，那些相邻的原子也只能采取特定的方向和这些原子相结合。比如水凝固后的晶体是冰，每一个水分子都有两个氢和一个氧，它们总是会和周围6个水分子相互靠近，规则地围成一圈。当这样的圈子越来越大时，冰就会出现六棱的特点。不过，从水结晶成完美的冰需要时间，我们看到冬天的雪花总是六瓣，就是高空中的水蒸气缓慢形成冰的结果。

当晶体熔化成液体的时候，因为原子的振动幅度加剧，电子跃迁扰乱了原子之间的纽带，于是这些原子相互之间的位置就会开始发生变化。如果这时候，原子（或分子）之间的吸引力足够大，大到还可以保持一个整体，那它们就会呈现流动的液体状态。否则，这些原子各自散开，它就变成了气态，二氧化碳便是这样，而这个过程就被称作升华。从这个过程，不难知道，物质能够在很宽的区间内保持液态，并没有那么容易。

非晶体的情况要复杂得多，因为各种原因，它没有能够形成“晶格”。比如玻璃，制造它的其中一个原材料就是石英，也就是二氧化硅的晶体。当石英熔化以后，会形成非常黏稠的液体，内部的硅原子和氧原子都会偏离原来的位置。此时，如果让它冷却下来，因为液体实在太过黏稠，原子无法回到原本的位置，这样一来，它就像是糖葫

芦下被糖包裹的山楂，保持原本的混乱状态被“冻”住了。如果石英凝固的时间足够长，它也可以像水变成雪花那样完美。在火山和海底，通常都可以找到由石英形成的上佳晶体，它们被称为水晶。而当石英被用来制造玻璃时，又有钠、钙这样的物质被加入了进去，原子的位置就更加难以归位，晶格再也不能形成。所以，当玻璃从流动的状态“凝固”时，我们并不能发现它除了流动性降低以外的变化，如果说它是流动性特别弱的液体，也有一定道理。

固体转化为液体虽然有很多情况，但液体转化为气体的过程却要简单很多。实际上，即便液体没有出现沸腾的状态，也还是会转化为气体，只是速度慢了很多，所以地面上的水会慢慢地变干。在这个过程中，液体最外围的那些原子，因为受到的吸引力要小于那些液体内部的原子，结果它们脱离了束缚，就变成了气体。如果对着液体持续加热，原子振动幅度加大，就连液体内部原子之间的吸引力都不足以将它们束缚，液体就会完全转化为气体。

对着气体继续加热，原子还会继续发生电子跃迁，直到这些电子就和“光电效应”中的电子那样，彻底和原子发生分离。这时候的物质，虽然还是气态，可是组成它的那些原子，却已经形成了各式各样的离子。因为这些离子中，正电荷与负电荷大致相等，因此这种状态也被称为“等离子态”。它是气体完全电离后形成的大量正离子和等量负离子所组成的一种聚集态。不停进行着核聚变的太阳，还有云层撞击出的闪电，都是天然的等离子态。不仅太阳，还有其他恒星中的气体也都处于等离子态。这种状态也有一些特别的性能，电焊时的高亮火光，就是一种人造的等离子态，它可以让钢铁瞬间熔化。

然而，不管怎么说，液态总是显得十分特殊：它和固态一样，原子之间会紧密地结合在一起，但是它同时又和气态一样，原子之间的位置可以错开，具有流动性。

所以，当物质处在液体中时，不同物质之间的能量交换也会最充分。在这颗星球上，海洋中存在着一系列洋流，温热的海水与冰冷的海水不断地交错流动，由此影响了各种气候，高纬度的北欧因此能够适宜居住，而南美洲则出现了海水与沙漠碰撞的场景。如果洋流停止或者倒转，温暖的地区可能会突然冰冻，而干燥的地区会迎来山洪，

这对于人类和很多生物而言，都将会是灭顶之灾。

我们在乎物质的不同状态，就是因为它们携带的能量。任何物质和能量之间，都有着不可分割的关系。在这个由物质和能量构成的超大系统中，没有任何物质会真正消失，能量也是如此。

然而，对人类而言，物质的意义远不止于此，我们还要将它们应用起来，在物质与人类之间形成互动。接下来，我们就来看看那些被人类熟练使用的物质。

# 6　万物争辉——物质是怎样为我们所用的

## 并不只是简单地混合——金属与合金

在元素周期表上，金属占据了118个元素中的94席。这其中，既有金、银、铜、铁、锡等人尽皆知的金属，也有像钌、铟、钽这样的罕见元素，还包括了诸如镉、锘、铊等一些人造元素。有着如此丰富的种类，金属元素注定会在物质科学中书写出浓重的一笔。

一个令人略有些意外的现象是，能够被广泛应用的金属，通常都不是某种纯粹的金属，而是制成一种被称为合金的物质。这种物质——合金就是由两种或多种化学元素（其中至少一种是金属）组成，如二元合金、三元合金和多元合金。它们同样具有金属的一些特性，却能改变纯金属性能的局限性，成为满足各种不同使用需求的优越性能的材料。

很长时间以来，金属影响了人类文明的发展。一般而言，进入青铜时代就是步入文明的标志，如青铜被大量用于铸造钱币，进入铁器时代的文明则开始走向成熟，至于影响深远的工业革命，更是由钢铁支撑起来的。到了晚期铁器时代，世界各地多已进入有文字记载的文明时代，铁器工具的使用排除了石器，并促进生产力快速的发展。这里所说的时代，通常指的是在考古学上的一个年代，如青铜时代一般指的是在考古学上继红铜时代后的一个时代。青铜就是红铜与锡的合金，故亦称锡青铜。中国在商代时期（约公元前1600年—约前1046年）已是高度发达的青铜时代，建立了冶炼青铜的工业。早在公元前3000年，美索不达米亚和埃及等地就已进入青铜时代。我国秦、汉以后，除青铜外，还出现一些其他的铜合金。最早出现的铜锌合金，即普通黄铜。黄铜就是铜锌合金的总称。后来又出现白铜，即铜镍合金。

尽管现代社会已经不再用某个金属来贴标签，但这并不意味着金

属不再重要。相反，更多新型的金属已经派上用场，铝、钛、镁等元素交相辉映，成为生活中不可或缺的金属材料，很难说到底是哪一种金属定义了新的时代。

当然，在金店里，我们的确还可以找到高纯度的黄金，其纯度即成色，一般以千分率表示。例如，“百足金”指的就是纯度（含金量）超过990‰的黄金，杂质不超过1%；“千足金”的纯度则超过999‰，以次类推。实际上，冶炼中不可能使其达到100%，因此，通常把纯度999.6‰以上的称为足金或足赤。然而，这些高纯度的黄金只是象征着财富，却并非理想的首饰材料。一方面，纯金只会显示出金色，难免有些单调；另一方面，更为要紧的是，纯金实在是太软了。在技艺高超的金匠手中，黄金首饰可以被打造成精美的镂空形态，可是戴上这样的首饰就得十分小心了，万一磕了、碰了都有可能发生变形，自然也就不那么好看了。

因此，为了使用起来更加顺手，黄金也常常会被制成合金。最初的分割熔合，可能只是为了降低每块金子的价值，方便交易——毕竟，米粒儿大的一颗小金珠就能换一大袋米，要是让它和其他普通金属熔在一起增加体积，就不会那么容易丢了。实际上，我们现在还会把纯金叫做24K金，就是这种方法的孑遗。古代进行黄金交易的人把金属中不同的组分称量出等重的24份，每一份都是一个Karat（这个词同样也被用在了其他珠宝的交易中，成为宝石的质量计量单位，并且演变成“克拉”。1克拉等于200毫克（1克拉等于205.3毫克是1913年前的旧制），其辅助单位是分，1克拉等于100分。为了避免混淆，代表黄金纯度的“karat”在英文中写作“carat”），其中有多少份是黄金，那么它就是多少K的黄金。24K金就是生活中的一般叫法，如18K的饰金就是纯度为18/24，即成色750‰。如果饰金的成色以“成”表示时，900‰的饰金就叫做九成金。

显然，这种办法将黄金分成了24个不同的纯度等级，数字越高则纯度越高。尽管这种“称金术”在如今早就不实用了，但是18K金或14K金却依然常见，它们通常是黄金与白银的合金。相比于纯金，它们的硬度更大，颜色也更多变，虽然价值打了折扣，但是制成的首饰还是颇受欢迎。

黄金是人类使用的第一种贵金属，世界很多地区都发现了早于当地文明诞生时期的黄金文物。这并非是一种巧合，只是源于物质的本性。

在太阳系形成之后的数十亿年里，地球也经历了无数次翻天覆地的变化。这里的“翻天覆地”并非是夸张——无论是气候环境还是地质结构，在地球上都从未有过须臾的平静，元素之间也在进行着激烈的碰撞。

我们已经知道，太阳系来源于一颗死亡的巨大恒星，那颗恒星以超新星爆发的形式释放出各式各样的元素，其中的一部分构成了地球的主体。早期的地球比现在更烫，到处都是流动的熔岩，这就意味着，密度更大的部分会因为引力的原因沉入底层。

通过现代技术对地球的结构进行探索，结果也的确如此：已经冷却的岩石覆盖在外表面，构成了地球的地壳，它虽然很薄，不足地球半径的1%，却是我们赖以生存的地方；仍然保持灼热的那些岩石形成了地幔，它们更像是一层受热软化的蜡烛，不停地蠕动，其中有一部分已经变成流动的岩浆，它也是地球主体的部分；科学家推测，至于铁、镍等更重的金属元素，就组成了地核，深入高压状态下的地球内部。

形象地说，地球就是一颗巨大的鸡蛋——薄薄的蛋壳，黏稠的蛋清，中间还有个鸡蛋黄。当然，我们还可以采取更精细的分析模式，把地球切分成很多同心球，就像洋葱那样剥开一层又一层，甚至给每一层都起一个名字，这在地质学上很有必要。但是在大多数时候，地壳、地幔、地核的划分就已经足够。再进一步的话，地核又可分为内核和外核两部分，外核深度约为2 900~5 100千米，推测为液态；内核深度约5 100千米以下至地心。据报道，1970年，苏联科学家超级钻探工程小组在地球上钻孔，垂直钻孔到达了12 262米深，成为地球上最深的钻孔。



地球洋葱模型

黄金的密度比铁大得多，它自然也会随着地球内部的运动堕入地核之中——以我们当今的技术，根本无力开采这些沉睡在地球内核的黄金。

所幸的是，地幔之中的那些熔岩十分黏稠，它们延缓了黄金沉降的过程。与此同时，元素周期表上排在第16的硫元素，在高温高压的作用下及时地与黄金结合，以硫化物的形式成为岩石的一部分。

灼热的地幔不停地蠕动，寻找着地壳的薄弱点，就像快要出壳的小鸡一样顶着地壳。倏忽之间，地球的某个地方山崩地裂，地震和火山纷至沓来，尘土冲上天空，岩浆滚落出来。正是在此过程中，黄金的硫化物也顺着岩浆来到地表。地表的压力骤降，黄金也与硫分离，成为游离态的金属，与岩浆冷却凝固后形成的岩石紧紧相抱。

又经过漫长的地质演变，昔日里坚硬的石头在雨雪风霜的摧残下变得松动，各种微生物以及苔藓野草也来凑热闹。最终，在这场被称为“风化”的漫长过程之后，岩石碎裂滚入河谷，又继续被磨成细小的砂石，夹杂在其中的黄金就这么留在了河滩之上。地球上主要的黄金产地大多位于河谷地带，长江上游被称为“金沙江”也并非是徒有虚名——这里的“金沙”的确很丰富。

大多数金属都没有黄金这么好的运气。比如铜，虽然也会和黄金一样经历从熔岩到地表的过程，但它和硫之间的结合力太强了，来到

地表之后并没有分离。甚至在经过漫长的风化之后，铜的硫化物也依然坚挺，需要通过一些手段才能转化为金属铜。所以直到今天，辉铜矿还都是冶炼铜的重要原料，它的主要成分是硫化亚铜（$Cu_2S$）。

铁和铜的经历类似，那些有幸没有落入地核的铁也以各种方式留在了地表上。自然界中丰富的黄铜矿，其主要成分被称作二硫化亚铁铜（$CuFeS_2$），实际上就是铁与铜的硫化物交织在一起。不过，除了硫以外，铁还有另一个好伙伴——氧元素。在风化过程中，空气中的氧气和铁结合，形成了更稳定的氧化物。中国南方的土壤呈现砖红色，正是因为土壤中含有大量的红色氧化铁（$Fe_2O_3$）。

相比于黄金，铜和铁都不能直接被人类利用，而是需要通过冶炼才能获得游离的金属。冶炼的原理并不复杂，只要将铜或铁从各自的矿石中剥离即可。然而，这需要能量，同时还需要一些成分带走矿石中诸如硫或氧这样的杂质。这样一来，实际操作就变得很有难度。人类掌握用火的技巧已有数十万年，但是炼铜的历史只有六七千年，冶铁的历史则更短。一般而言，冶铁技术发明于原始社会的末期，它标志着冶金史上进入新阶段。人类锻造铁器的起点也就在公元前1400年左右，我国在春秋晚期（公元前5世纪），大部分地区已使用铁器。

不过，和黄金相仿的是，为了提高铜和铁的性能，人们通常也要把它们加工为成合金。

铜的合金品种很多，古人就已经发明出青铜和黄铜，它们分别是铜混合了锡（或铅）和锌的结果。古代中国人还发明出一种铜和镍的白铜合金，看起来就和银子差不多，至今还被用来制造钱币。

铁最出名的合金就是钢，它是由铁和碳形成的，其中碳的质量分数在0.025%～2.06%。如果含碳量更高，它就被称为生铁。生铁不容易变形，但容易开裂；如果含碳量更低，它又会被称为熟铁，实际上已接近于纯铁，质地软得跟皮带一样。所以，铁通常都会被加工成钢再使用。而在现代技术的加持下，钢的种类也越来越多，比如常用于机械的锰钢，可以用作防弹甲板的钨钢，还有不容易生锈的不锈钢，等等。

还有更多的金属元素呢？它们的命运甚至还不如铁和铜这般顺

利。

比如铝，它是地壳中含量最大的金属元素，经过漫长的演变，这种元素绝大多数都和氧元素结合在一起，形成被称为“铝土”的矿物（$Al_2O_3$）。此前我们已经知道，铝和氧之间的结合力非常强，所以想要把铝从矿石中提炼出来，万分困难。古人用炼铜或炼铁的方法，根本提炼不出铝，直到电被发明出来并广泛使用以后，才有了电解炼铝的工艺。即便如此，因为矿石超强的结合力，它的熔点实在太高，故而还需要在其中加入一种助熔剂——顾名思义，这就是为了帮助矿石熔化。这种助熔剂被称为冰晶石，就因为它可以起到降低熔点的作用而得名，其主要成分是六氟合铝酸钠（$Na_3AlF_6$）。

铝也不是最难冶炼的金属。

在元素周期表的下方，通常还会多出两行，它们分别被称作镧系和锕系。它们本该排在元素周期表的第三列，但是这样会让表格显得太长，故而一般的印刷版本都会将它们截到最下方。

锕系元素大多数是人造元素，在地球上的存量极低，只有为数不多具备开采价值的元素，例如钍和铀，它们主要都被用在了核电厂中。

镧系元素可不一样，它所包含的15种元素，连同周期表上第三列已有的钪和钇，合起来被称为稀土金属。这些金属元素个个身怀绝技，可以被应用在很多高科技设备中。比如有一种叫钕的元素，它就可以被用来制造强磁铁。所以，稀土元素也常被称作“工业维生素”，用量不多却不可或缺。

然而，冶炼稀土元素可不容易。它们不只是会像铝那样，其矿石具有很高的熔点，而且，这些元素的性质实在是太相似了，想要把它们分离出来，就好比从长得一样的多胞胎中找出其中一个，那可是相当不容易。直到现在，能够掌握全套分离技术的国家也寥寥无几。中国有一位科学家叫徐光宪（1920年—2015年），很早就看到了稀土元素的巨大价值，也正是在他的领导和呼吁下，中国的稀土提炼技术如今已经在全世界领先，有人把他誉为中国的“稀土之父”。

从黄金到稀土，人类花了好几千年的时间，也还是没能把金属物

质的世界研究透，大部分金属，我们还都没有找到它最合适的用途，这还有待于我们继续努力开发。

而在元素周期表的另一部分，也就是非金属元素，虽然成员的数量不多，但是它们构成的物质却在地壳中占据了很大的比例，人类对它们的利用也从未停歇过。

## 从石器到陶瓷

人类的历史是从石器时代开始书写的——漫山遍野的岩石提供了最初的开发工具。石器时代是考古学上人类历史的最初阶段，属于原始社会时期。那时，石器是人类劳作的主要工具。

岩石星球并不总是会布满岩石。

月球和地球形成的时间相似，也同属于岩石星球。然而，月球比地球小得多，它的引力不足以在月球表面支撑起大气层。于是，当月球被太阳照耀到的时候，温度可以高达100多摄氏度，可是等到月球上进入黑夜时，温度又会低达零下100多摄氏度。中国的探月专家认为，巨大的温差，让月球上的岩石不停地发生着膨胀与收缩，它们会比地球上的岩石更容易风化崩解。而且，月球上的火山也已经偃旗息鼓，它也很难从月球内核补充新的岩石。因此，月球的表面绝大部分都被细碎的小石子或石粉覆盖。当“嫦娥”探测器登陆到月球的时候，深深的辙印很清晰地证明了这一点。

地球上的岩石绝大部分都含有硅和氧这两种元素，它们在地壳中的总含量超过了75%。被人类用来打造石器的原料通常也是以这两个元素为主体——很大一部分原因也是他们别无选择。

如果岩石中只有硅和氧，那它就会被称为石英。石英的化学成分就是二氧化硅（$SiO_2$）。在河沙中，石英就占了很大的比例，而我们此前也已经提到，石英形成的完美晶体便是水晶。不过，大多数石英都会和其他一些元素依靠化学键组合在一起，它们形成的这类物质被统称为硅酸盐。

按照成因，地球表面的岩石通常会被分为三类：火成岩、沉积岩

和变质岩。火成岩来自火山喷发之后冷却的岩浆，沉积岩通常是河底小碎石被挤压在一起形成的大块石头，变质岩则是地下高温高压下形成的岩石。它们可以相互转化，例如火成岩风化后会进入河道形成沉积岩，而火成岩与沉积岩都可以在地下发生“变质”，形成变质岩。

不过，对于远古的人类来说，这些岩石的来历并不重要，他们更在意的是，这些岩石是不是能够满足需要——主要是狩猎和日常的穴居生活。

幸运的是，硅酸盐虽然在地球上随处可见，相当普通而易得，但它的性能却很卓越。硅和氧之间的化学键具有无限延伸的能力，所以，当我们抓起一块硅酸盐岩石的时候，它内部的原子全都彼此连接，这就让它具备了可观的硬度与强度。或许将原子编织成手掌大的一块石头并不让人吃惊，但是如果知道“艾尔斯巨岩”的话，一定就不会再这么想了——这块巨大的石头位于澳大利亚，因为含铁量高而通体呈现红色，绕着它走一圈需要好几个小时，石头露出地面最高的部分超过埃及胡夫金字塔的两倍——后者由数万名劳工用石条堆砌了大约20年才完工。

不难看出，石头有着超凡的承压能力，原始人很自然地就发现了这一点，用石头制作各种工具，大到石斧，小到石簇（箭头），显著提升了他们的生存能力，也构建出史前伟大的石器时代。

然而，石器也有个严重的缺陷——它太硬了，加工起来很困难，大多数时候只能靠不同的石头相互摩擦才能造出合适的造型——这很费时间。

还好地球上的硅酸盐并没有让我们失望。那些细小的岩石落入河谷后，在水流和石块彼此的撞击之下，变成细小的河沙。生物的出现，又让河沙中多了很多有机质——我们很快就会说到它——最终变成河底的淤泥，逐渐被河水推到了岸上。相比于河沙，淤泥中的硅酸盐颗粒已经细小了很多，但它并没有能够和这些有机质很好地融合。它的确很肥沃，却又难以保留肥力，大多数植物并不能直接在淤泥上生长。一旦这些淤泥在阳光下暴晒几天，它们就会出现龟裂，最终形成扬尘。

又经过多年的风化，这些淤泥中的硅酸盐还会继续瓦解，变成肉眼无法分辨的小颗粒。这种极小的颗粒具有很强的吸附力，它可以和有机质均匀地融合在一起，形成我们更熟悉的土壤——有一些土壤还颇有黏性，故而也被称为黏土。

在地球上，黏土虽然不及岩石那么普遍，但也还算常见。也不知道从什么时候开始，人类学会了用火炙烤黏土，发现黏土在被烤制一段时间后，居然会发生硬化，变得像石头一样。这种材料，我们现在称之为陶。

烧陶之前，人们会将它们捏成特定的形状，加热之后就可以制成饭碗、酒杯这样的造型。相比于石器，陶器在制成这类工具方面有着无以比拟的优势。于是，陶器也迅速在地球上各个部落流行开来，成为人类文明早期最重要的材料。

在中国商朝时期，人们在烧陶的过程中，又发现了一种新的材质。可能是因为机缘巧合，有一些陶器中出现了钙元素，它在硅酸盐中的作用，就如同是冰晶石在铝矿石中一样，可以起到助熔的效果。至于这个配方，又有点像是玻璃。于是，陶器表面有一些硅酸盐发生熔化，冷却后就变成晶莹剔透的玻璃状硅酸盐，所形成的连续玻璃质薄层被称为“釉”。又因为这种釉中含有少量铁元素，在灼烧之后会呈现青绿色，故而又被称为“青釉”。青釉器就是瓷器的开端，它启发了中国人开发高岭土这种硅酸盐黏土烧制瓷器的工艺，又经过两千多年的发展，产生了青瓷、白瓷、青花瓷等各式精美瓷器，成为丝路驼队还有远洋商船上的货物，远销全世界。



西汉时期的青釉器

可以说，在人类社会如此悠远的历史中扮演着如此重要的地位，陶瓷是独一无二的。而在现代，陶瓷又在高科技产品中占据着特殊的地位。

不同于古代的陶瓷，现代陶瓷并不只是硅酸盐材质，还包括硼、碳、氮、铝、磷、硫等多种元素，其中大部分是非金属元素。它们应用的领域也远不止于锅碗瓢盆这样的日用品，还有诸如电绝缘体、磁性体等各种材料。

有一些新型陶瓷的性能，甚至颠覆了我们过去对于物质的认识。金刚石是天然存在的最硬物质，它是完全由碳原子拼搭起来的物质，每一个原子都和相邻的四个原子之间形成了化学键，自然也就十分牢固。一直以来，人们都相信，它会保持最硬物质的记录。然而就在2013年，中国有一位叫田永君的科学家，他领导的团队却合成出了一种氮化硼陶瓷材料，比金刚石还要硬，又一次刷新了人们对物质的认识。

陶瓷究竟还有多少奥秘，我们现在还不能完全参透。它依然是我们生活中不可缺少的一部分，但也会时常出现新的陶瓷材料，带来新的惊喜。

陶瓷虽然多变，可是在所有种类的物质中，即便是陶瓷加上金属，也不及物质种类的1%，剩下的绝大部分物质都属于有机物。

## 无穷无尽的有机物

有机化合物简称有机物，就是含碳化合物的总称。只有几种物质例外：通常二氧化碳、一氧化碳之类的简单含碳化合物会被认为是无机物，但它们和有机物的关系也非常紧密。

在元素周期表上，碳元素排在第六位，看起来平平无奇。然而，科学家们却很早就注意到它了。

碳元素第一次引起学术界的震撼，来自拉瓦锡完成的一项疯狂实

验：烧一颗钻石，看看那样会产生什么。金刚石经过人工琢磨后的产品就是钻石，因为它实在太硬了，不能被切削，所以在当时并没有人知道它究竟是什么物质。而且，钻石也太珍贵了，一般人也没有经费去研究它，但是拉瓦锡本人是法国贵族，他有这个财力。

实验的结果出乎意料——钻石中的化学成分只含有碳元素。

碳元素也是煤炭中最主要的元素，它和包括氢、氧、硫在内的很多元素形成了各式各样的分子。如果把煤炭中的其他元素全部脱除，只剩下碳元素，那么最终得到的就是石墨，它黑黝黝的外观，看起来和钻石完全不相干。

然而物质世界就是这样，钻石和煤炭居然如此相似。

后来，盖-吕萨克的实验室里来了一位名叫尤斯蒂斯·冯·李比希（Justus von Liebig，1803年—1873年）的年轻学者。此时，盖-吕萨克和道尔顿的论战还没有结束，“分子”的概念也还在角落里坐着冷板凳，没什么人在意。1830年，李比希在前人工作的基础上，使碳氢分析发展成为精确的定量分析技术，他也成为德国历史上非常重要的化学家。

早在1815年，印度尼西亚的坦博拉火山爆发，成为人类有记录以来最大的一次火山爆发，至今还保持着记录。火山喷出的烟尘实在是太厚重了，长年飘在天空中，甚至在第二年，包括欧洲在内很多地区没能迎来夏天，因为阳光被空气中的火山灰吸收了。不仅如此，当云层转变为雨水时，火山灰中的很多物质也会溶解在雨水中，特别是二氧化硫这样的物质会转化为硫酸，于是雨水就成了破坏性很强的“酸雨”。

在各种因素的叠加之下，全世界都在1816年遭遇了不同程度的粮食减产，有些地区甚至出现了灾荒，仅欧洲就有数十万人的死亡和这场灾难相关，部分国家因此陷入动乱。少年时期的李比希目睹了这场人间惨剧，这也促成了他一生中最关心的工作——研究如何让粮食增产。他成为农业专家，还发明出最早的化学肥料。他经过研究发现，正是因为在土壤中缺乏了一些特定元素，粮食才不能很好地生长。其中，植物最容易缺失的三种元素是氮、磷、钾，因此最流行的化肥就

以这三种元素为主。因为它们的元素符号分别是N、P和K，所以这类化肥就被为NPK肥料。

然而，李比希还注意到，土壤中的碳元素似乎也不可小觑。

在此之前，英国科学家普利斯特里已经发现了光合作用。他是氧气的发现者之一，设计出的实验曾经启发了拉瓦锡。植物在进行光合作用时，会吸收空气中的二氧化碳和水，然后转化为葡萄糖，植物会以葡萄糖为原料，加工出它需要的各种物质。

光合作用解释了更早时候的“柳树实验”。17世纪时，比利时（当时还叫尼德兰）科学家巴普蒂斯塔·范·海尔蒙特（Baptista van Helmont，1577年—1644年）为了弄清楚植物的养分从何而来，在大花盆里种下一棵柳树。五年后，这棵柳树的重量已经和成年人相仿，但是土壤减少的重量却只相当于两个鸡蛋而已。尽管当时还没有人能够证实“质量守恒定律”，但是海尔蒙特还是合理地推测，柳树生长时所需要的各种成分，主要来自于空气，普利斯特里最终解释了这个原理。

既然光合作用说明植物中的碳元素是植物吸收了空气中的二氧化碳才形成的，是不是土壤中的碳元素就没什么用了呢？李比希通过实验证明，土壤中的碳元素虽然不多，可它对于植物而言，甚至比其他元素更重要。就是在这些现象的启发下，李比希提出设想，认为含碳的物质是生命体需要的，是它们让生物变得生机勃勃，故而被称为有机物，与之相对，不含碳的物质便是无机物了。

李比希找出生命体和碳元素的关系，并由此归纳了“有机物”的范畴，后世便尊他为“有机化学之父”。但他不只是在农业和有机化学方面有点造诣，同时还是一名教育家，非常善于将自己的思想传达出去。德国著名的有机化学家凯库勒（Kekule，1829年—1896年）还是学生的时候，就听说李比希的讲座很有趣，去听了一次之后，就迷上了有机化学，并投入到李比希的门下。

这时候，“化合价”这个概念也已经被提了出来，凯库勒便开始用化合价的概念去解释有机物为什么与众不同。他首先确定，在有机物中，碳原子总是倾向于形成四价，最多可以同时和四个原子结合，而

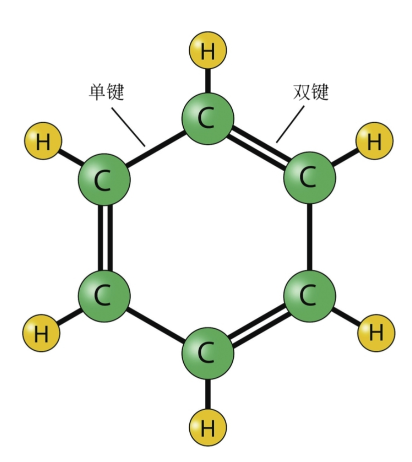
且碳原子和碳原子之间也可以互相连接，这就构成了有机化学最核心的基础。

后来，“分子”的概念也被科学界承认了，凯库勒就更进一步，确认了很多有机物的结构。其中最著名的莫过于“苯”，至今在教科书上还流传着他的传说。

苯分子有6个碳原子和6个氢原子，按照当时的分子理论，虽然可以绘制出一些不同的结构，可是这些设想中，却没有哪一个结构是合理的。凯库勒对这个问题也是百思不得其解，白天研究，连夜里都没闲着。有一天，他做了个梦，梦到一条蛇回头咬住了自己的尾巴，受此启发终于想通，苯可能是一种“环状结构”。后来，又有人设想出苯环结构的其他形式，因此凯库勒绘制出来的结构就被称为“凯库勒式”，以示区分。



凯库勒的衔尾蛇




苯分子的凯库勒式

尽管这个故事有一些附会的成分，但是“凯库勒式”的出现，的确颠覆了人们对于有机物的想象。碳原子的化合价为四价，虽然并不是最高的，但是碳原子之间却几乎可以无限连接，而且它们还可以形成

环状、笼状、树枝状等各种结构，这就让有机物的形式变得非常复杂。

不仅如此，很多有机物还存在同分异构体。就像凯库勒研究苯的时候，最初设想的那些分子结构，后来有一些也通过实验被发现了。它们虽然也有6个碳和6个氢，却和苯分子有着截然不同的特性。这些原子组成化合物的分子式相同，但具有不同的结构和性质，就被称为同分异构体。这在有机化学中极为普遍。

于是，碳原子的连接千变万化，含碳的分子也难以计数。到现在为止，人类发现的物质有上亿种，其中无机物不过十余万种，有机物占了绝大多数。

凯库勒揭示了碳原子的结合规律，也有助于人们厘清金刚石与石墨的关系。在金刚石中，每一个碳原子都和另外四个碳原子相结合，它们形成了立体结构，所有的电子都参与形成了化学键，每个碳原子的位置都保持稳定，不会发生位移，所以金刚石的硬度非常大。至于碳原子形成的石墨，则是每个碳原子与另外三个碳原子以平面的方式结合，这样一来，四价的碳原子就留出了一个自由的电子，所以石墨就可以靠着这些电子自由地传递电流，不像金刚石那样是绝缘体。

与此同时，有些物质虽然不含碳，但是它们居然也会采取有机物那样的方式构成分子，比如元素周期表上排在碳元素之前的硼元素，还有排在碳元素下方的硅元素。这些元素在和氢元素结合的时候，有时也会遵循和碳元素相似的规律，因此科学家们对它们也充满了兴趣。物质之间的“有机”组合，似乎还给物质赋予了生命，我们会在下一章谈到这一点。



石墨烯

换句话说，有机物虽是含碳的物质，但是“有机”的内涵却丰富了许多。在我们的生活之中，“有机食品”正在流行，它们是在“有机农

业”生产体系下，生长于良好的自然生态环境，无添加剂、无污染、纯天然等按有机农业生产要求，以及相应标准加工生产出来的一切农副产品。它们的确也都富含有机物，然而这并非是它们被认证为“有机”的本质原因。

之所以我们会对不同物质的认识越来越深，是因为我们现在已经有了越来越多的分析手段，可以看到物质的结构——就像苯分子，在电子显微镜下就可以直接看到它那优雅的六边形环状结构。正是多变的物质结构，才让物质世界变得如此丰富多彩。

## 结构决定性质

金属与合金、陶与瓷，还有石墨与钻石，有机物中的同分异构体，它们彼此之间的联系与差异，全都说明了物质世界中最重要的原理之一：结构决定性质。

对于很多物质而言，它的组成固然很重要，但也只有弄清其中的各种元素如何组合在一起，才能知晓它会有怎样的特性。

2020年年初，COVID-19病毒开始在全世界流行，成为一种让人谈之色变的瘟疫。这种病毒属于冠状病毒，而且还是过去从未被发现过的病毒种类，故而也被称为“新型冠状病毒”。

为了应对突如其来的疫情，物资方面的准备显得尤为重要。除了口罩、手套以外，快速检测出病毒的试剂也不可或缺。这种试剂最快可以当场确定一个人是否已经被病毒感染，这样就能更有效地进行防疫了。

检测病毒的原理倒也不是很复杂。

我们人类的身体中有一套防御体系，它被称为“免疫系统”，负责对外来的物质进行戒备并应对处置。当病毒成功地感染了某个人之后，它就会开始复制，在人体内的数量越来越多。病毒释放出某些被称为抗原的物质，这会引起免疫系统的警觉，发现外敌入侵。作为回应，免疫系统也会释放出一种抗体，用来对抗病毒。所以，病毒检测试剂只要能够测出这些抗原或抗体是否存在，就能证明一个人是不是

被病毒感染了。

当然，这些都是间接测试的结果，我们还可以使用直接检测的手段。病毒中含有一些叫“核酸”的遗传物质，它们的结构特殊，只要确认这些核酸是否存在，同样可以证明受检者是否被病毒感染。如此检测的结果更准确，但检测时间相对更久一些。

在这些不同的COVID-19检测方法中，抗原检测的方法最简单，它只需要从鼻腔或咽部取样，再将样本简单处理后滴到试纸上。几分钟后，如果试纸上只出现一条线，就说明受检者是阴性，没有被病毒感染，如果出现两条线，就说明受检者是阳性，已经被病毒感染。

这种用两条线来标记出阴性或阳性的方法，就是现在最流行的“胶体金”法。

在“胶体金”中，黄金被磨成了极其微小的粉末——小到无法用肉眼看清。它们的颗粒大小通常不足一微米，进入到纳米的尺度，故而也被称作金纳米颗粒。1纳米只有1毫米的100万分之一，所以这些金颗粒还不及头发丝的百分之一粗。这时候的金，已经不能再被称呼为“黄金”。它们可能会呈现出粉红色，但也可能会是酒红色或紫红色，这取决于它们的实际大小。

金纳米颗粒的吸附性很强，像是黏胶一般，遇到抗原就会相互吸引。这并不只是金才具备的特性，大部分固体物质在这个尺寸下都是如此，“胶体”的名称也就由此而来。

不过，胶体金的特色在于，它在吸附了抗原后，就会呈现出很有特征的颜色。这时候，只要在试纸上埋上一道特定的抗体，因为抗原会和这种抗体结合，那么当胶体金附着了抗原沿试纸通过时，就会把它们拦截下来，形成一道由胶体金显现出线条。




胶体金抗原检测试纸

实际测试时，试纸上会有两条线，分别被称为控制线和检测线。在处理受检样本时，还会加入另外一种已知的抗原，它附着在胶体金上，就会和控制线上的抗体结合，显现出颜色。如果控制线不能显色，就说明这张试纸已经失效了。如果样品中含有需要检测的COVID-19病毒，那么检测线上也会显现出颜色，这就是阳性样本会出现两条线的原因。

黄金是人类最早学会利用的金属，然而同样是金元素，既可以用来打造首饰，可以被卢瑟福用来完成金箔实验，还能被做成金纳米颗粒用在病毒检测中。不同的用途，彰显的是金元素多彩的性质。

事实上，每一种物质都是如此。

而在这些不同特性的背后，就是物质不同的结构。物质的结构决定了它的性质，于是我们可以通过调整物质的结构，实现它们特殊的功能。

整个自然界遵循的也都是同样的规律，这并不会以人的意志为转移。甚至就连我们自己，也包括所有的生命，都要依靠物质去实现生理功能——根本上说，是物质构成了生命。

从物质到生命，这是一个复杂的工程，我们将在下一章中窥一窥其皮毛。

# 7 活着的奇迹——赋予生命的物质

## 从简单到复杂

物质为什么能够演变出生命？这可能是物质科学需要面对的终极问题。这是因为，生命体和非生命体之间显然存在着某种不可逾越的界限，我们并不知道物质世界怎样打通了这个界限。

很长时间以来，科学界认为存在着两个物质世界，也就是有机物世界和与之相对的无机物世界，并且这两个世界的物质之间并不能相互转化。比如我们赖以生存的葡萄糖，它是有机物，就只能从动物或者植物这样的一些有机物中获取，不能靠无机物转化而来。这样的分类方式，认为生命和非生命的界限天然而永久地存在，从而彻底回避了上述问题。

不过，这样的处理某种程度上也只是掩耳盗铃，它引发了人们对更多问题的质疑。我们能够从植物中摄取葡萄糖，那么植物又该从何处摄取它呢？又或者，有机物构成的木头在燃烧之后变成了无机物，如果无机物不能转化为有机物，岂不是有机物会越来越少？

经过数年的研究，德国化学家弗里德里希·维勒（Friedrich Wohler，1800年—1882年）于1828年首次以无机物氰酸氨中原料合成出尿素这种有机物，对于这个问题的讨论才步入正轨。这项工作打破了有机物和无机物之间的人为界限，动摇了当时盛行的“生命力学说”，指出了有机物的合成方向。时至今日，“有机物”已不再是本意所指的“有生命力的物质”，而是指绝大多数含碳化合物，只是为了方便研究才沿用至今。

如此一来，有关生命起源——从无生命物质形成原始生物体的过程——的问题，似乎也有了寻找答案的方向。那些不具有生命特征的物质，很可能就是生命最初诞生的地方。

实际上，在自然条件的撮合之下，新物质的生成速度超乎想象。

太阳系诞生之时，刚刚形成的地球还是一颗灼热的岩石球，在它的表面，连水都是蒸汽的状态，与氢气、氨气、甲烷等各类分子构成了地球早期的大气。然而，事情很快就发生了变化，剧烈的地质运动又释放出二氧化硫、二氧化碳等物质，让本就活跃的大气层变得更加热闹。它们彼此之间会发生一些简单的反应，如今空气中最常见的氮气便在这个时候形成。

还有一些反应会引起我们格外的注意。

正如日本隼鸟探测器所证实的那样，地外小行星上可以找到氨基酸分子。事实上，即便是在太阳系以外的空间，氨基酸都是很常见的一类物质，它们是构成生命体的核心组件之一。

所有的氨基酸分子都含有碳、氢、氧、氮这四种元素，例如最简单的氨基酸叫做甘氨酸，分子中只有2个碳原子、2个氧原子、1个氮原子以及5个氢原子。当氨气、甲烷和水共存时，就已经凑足了这4种元素，在闪电的作用下，可以形成包括甘氨酸在内的一些简单氨基酸，这一反应过程也早已被实验证明。

而当大气中出现二氧化硫后，体系变得更加复杂，由此形成更多种类的氨基酸。有一种被称为胱氨酸的氨基酸，其中就含有硫元素，它在搭建生命的过程中居功甚伟，后面还会提到它。

可见，早在地球形成初期，以氨基酸为代表的有机物就已经大量出现。随着地球表面的温度逐渐降低，海洋开始形成。这些能够溶解在水中的氨基酸，在海水中逐渐富集。



富集氨基酸的海水就如同肉汤

氨基酸有着非常特殊的结构：它的一边被称为氨基，另一边则被称为羧酸基团（羧基），这也是“氨基酸”名称的由来。氨基和羧基可以发生化学反应，结合起来。如此一来，当两个氨基酸相互靠近时，它们就像两只锁扣一样连接在一起，形成一种被称为“肽”的新物质。在肽的两端，分别残留了一个氨基与一个羧基，因此它还可以继续连接其他氨基酸。以此类推，肽会变得越来越长——这种特别长的肽就被称为蛋白质。

正所谓“海纳百川”，海洋中的几乎可以找到所有常见的元素，化学反应的类型也比大气之中更复杂。特别是磷元素的加入，在有机物和无机物之间架起了一道桥梁。水中的磷元素很容易和氧形成一种被称为磷酸根的无机离子，它有三个接口，每一个接口都可以连接一种物质，其中也包括磷酸根本身。

如果磷酸根的其中两个端口都和其他磷酸根彼此连接，如此延续下去，就可以形成一条由磷酸根串起的长链。与此同时，每个磷酸根还留出一个空闲的端口，如果这个端口迎来一种叫核苷酸的有机分子，那么这条由磷酸根连接而成的“项链”就成了核酸分子了。它是让生命延续下去的重要奥秘，后面也会再提起它。

如果磷酸根青睐的是钙离子，那么结果又会很不一样。由磷酸根和钙离子形成的物质被称为磷酸钙，它是一种非常坚硬的物质，就和石头相仿。经过氢氧根离子的修饰后，磷酸钙会转化为“羟磷灰石”，其硬度刚好能够支撑起生物的重量。如今，我们在地球上发现的所有

脊椎动物，骨骼都是以羟磷灰石为主体。

甘油是另一种带有三个端口的物质，学名丙三醇，是一种有机物，也是生命的重要组成部分。甘油被称为“油”，也出现在很多油脂之中，但它本身却不是油。它的结构也不复杂，即便是地球诞生不久的恶劣环境下，也不难形成。

不同于磷酸根，甘油之间很难彼此相连，却几乎专一地与羧酸类物质连接，由此形成一类被称为甘油三酯的物质，分子结构就如同三叉的烛架插满了蜡烛。甘油三酯富含能量，几乎所有的生物体都靠它储存能量，人类也不例外。只不过，对衣食无忧的现代人来说，身体中过多的甘油三酯已经成为负担，体检时通过验血判断的“血脂”指标，指的就是血液中的甘油三酯浓度。

甘油和磷酸根还可以相互配合，与羧酸形成一类被称为“磷脂”的物质。磷脂的性质很特别，它的一端亲和水分子而排斥油脂分子，另一端却又亲和油脂分子而排斥水分子，因而被称为双亲分子。在生命体中，磷脂分子也发挥着举足轻重的作用，特别是卵磷脂，能够利用自身的双亲特性，调节身体中油脂的新陈代谢。不止如此，由两层磷脂构成的细胞膜，将细胞内外隔开，允许营养成分从细胞外向内流动，代谢产生的废弃物又能由内而外流动，细胞内部的化学反应得以顺利地进行。

近代科学研究说明，生物只能通过物质运动变化，由简单到复杂逐步发展形成。随着时间的推移，地球上的分子种类越来越多，结构也越来越复杂，生命体必需的物质基础也逐渐完善。然而，第一个生命到底是什么时间出现在地球上，这个问题并不容易回答。这不只是要考虑生命需要哪些物质，更要研究地球自身的环境。

地球拥有适中的温度，海洋可以因此保持液态，物质可以因此更容易地发生反应。不过，温度的作用不止于此：温度的高低决定了化学反应的速度。如果地球的温度太低，化学反应的速度很慢，那么生命物质形成的速度也很慢，不足以支撑生命的需求。反之，如果温度太高，反应速度太快，那么好不容易形成的各种生命物质，也会因为发生化学反应而衰减，难以积累到足以出现生命的程度。

即便已经万事俱备，生命形成的概率也并非是百分之百，还需要依赖“涌现”（emerge）这一过程。当若干简单的部分叠加在一起时，会形成更复杂的系统，然而复杂系统的某些特性，却是简单部分原先所不具备的，这种“整体大于部分之和”的现象便被称为涌现。我们每天都会和涌现现象打交道——就像破折号前的这句话，如果说成“现象道交打们我涌现和会都天每”，没有人可以轻松解读它的含义，尽管所有的字都还在，只是顺序被打乱了而已。由此可见，这句话的含义并非是由所有字原本的意思叠加而成，它超出了字意的总和。或者说，只有这些字以恰当的方式组合在一起，才能“涌现”出最终的含义。

与之类似，当所有生命要素全都凑在一起时，也只有采取特定的方式，才有可能涌现出生命。这个过程有着太多的不确定性，就像掷骰子那样难以预测。可喜的是，我们的地球成功了，最终打破了生命与非生命的界限，这也是物质世界的神奇之处。

但我们是否能够进一步知道，生命究竟是什么？

## 低熵系统

想象一下，你正在火锅店里大快朵颐，刚刚煮熟的虾滑在红汤中翻滚——你已经注意它很久了。面对这种并不太适合用筷子夹起来的食物，你会娴熟地拿起漏勺，将它从汤中捞起。从漏勺孔洞中流淌下的火锅汤料，还夹杂着几颗花椒。至于那些滑溜溜的宽粉，通常也只会顺着漏勺的边缘滑落到汤里。因此，你会很顺利地捞出你想要的虾滑。

在生活中，我们会借助类似于漏勺这样的很多工具进行物质的分离，而且多数时候都随意得有些漫不经心。把被子晾晒到栏杆上时，随手抄起一只衣架朝着它拍打几下，深藏在被褥中的灰尘与螨虫，就会在机械力的作用下喷溅而出；雨天里被水打湿的课本，放在取暖器旁烤一烤，纸上吸附的水分便消散了；喝茶的时候，轻轻地吹一吹，就可以避免把茶叶沫子喝进嘴里……衣架、取暖器，甚至只是我们呼出的气体，都可以被我们用来分离不同的物质。

生活中的这些动作，我们操作起来之所以会如此熟练，是因为它们其实是生命的本能。任何一种生命体都必须“学”会分离物质：鱼鳃的结构是为了分离水中的氧气，这样它们才能在水中呼吸；树木强大的根系将水和矿物质从土壤中分离出来，顺着树干送到每一片树叶；培养皿中的细菌会从营养液中分离出它们需要的养分，几乎每半个小时就可以繁殖出下一代……

换一个角度而言，生命无非就是将合适的物质搬运到合适的位置，我们人类也不例外。所有能够供我们祖先生存下去的活动，无论是采摘野果还是追踪猎物，都是从自然界中获取那些我们需要的物质，舍弃掉不需要的物质。

当我们的生存能力已经足够强大，能够在一定程度上改造自然界之时，所做的事情无非还是在分离各种物质：从水稻之中，我们收获了米粒，以此作为最重要的主食之一；从黏土之中，我们剥离了水和有机质，使之形成坚硬的硅酸盐并成为陶瓷；从铁矿石中，我们脱除了矿渣和氧元素，由此锻造出钢铁……

靠着分离物质的各种方法，包括人类在内的生命体才活了下来，一代又一代地繁衍，并将生存的奥秘传了下来。直到现在，我们还在做着相同的事情，而且就和火锅里捞虾滑一样寻常而频繁。

这些分离物质的奥秘，科学上用一个被称为“熵”的热力学系统状态的物理量进行描述，它还有一个更接地气的名称：混乱度。换言之，有序而规则的物质，其中的“熵”较低；混乱而无规则的物质，其中的“熵”相对更高。

相比于其他物质，有序是生命体最为直观的特征之一，也就是说，生命体中蕴含更低的“熵”，其结构和运动状态具有确定性和规则性。这一概念，量子物理学家薛定谔曾在他那本著名的《生命是什么》中进行了风趣而详细的阐述，但是想要理解它，并不需要太高深的物理知识。

试想你现在手上有一杯水，你用力将水朝斜上方泼了出去，一两秒钟后，这些水就会在重力的作用下落在地面上。每一次泼水，你都能得到一个随机的图案，每一次都不相同。那么，有没有一种可能，

你随手泼出去的水，水迹图案会是人的形状呢？



随意泼出去的水会洒出怎样的图案？

这种可能当然存在，如果你愿意试上很多次的话——这里的“很多”通常会是一个巨大的数字，“亿”这样的单位和它比起来，也只能说是微不足道。

人体中超过2/3的部分都是水。如果我们将这些水收集在更大的杯子里，随便泼出去，这些水重新按照人体中这样分布的概率，几乎就是零。

出现这种情况，就是熵在“作祟”。任何闭合系统都会自发向着混乱度增加的方向发展，也就是说，任何物质总体上都会变得越来越混乱。这一定律被称为热力学第二定律，不过人们更习惯以“熵增加原理”（也叫熵增定律）来称呼它，作为热力学第二定律的定量表述。

水是由一个个水分子构成，泼出去的水，混乱度增加，水会以更随意的方式铺展开，很难形成我们希望形成的图案。也有一些画家爱上了这种不确定性，他们会用喷射颜料的方式作画，那些不受控制形成的随机图案，恰恰成为最富有个人特征的作品。

然而，人体中并不只有水，还有成千上万有机物与矿物质，这就让“有序”变得更加难能可贵。事实上，别说是这么多种物质，哪怕只是在水中添加一点食盐，系统的混乱度都会呈指数级的增长。

对生命起源的探究发现，是地球早期的海洋，孕育出最初的生命体，人类也不例外。因此，人体中的各种体液，大多还保留着海洋的特征，尤其是胎儿生存所仰仗的羊水，羊水中富含的氯化钠也正是海

水中最主要的盐分。羊水存在于孕妇的子宫之中，实质上是由各种盐分混合而成的水溶液。胎儿在这样的环境中慢慢长大，就像是我们的远祖在海洋中生活那样。

地球表面覆盖着辽阔的海洋，受制于不同的地理环境与气候条件，不同海域的盐度有着很大的区别。例如地中海，它被大片陆地环绕，海水的蒸发量高，盐度也相对更高；与它仅仅通过直布罗陀海峡相连的大西洋，因为接纳了无数河流的淡水，其盐度就要更低一些。于是，在狭窄的直布罗陀海峡，就会出现非常奇妙的现象：地中海底部的浓盐水会朝着大西洋方向流动，而在表面，大西洋的稀盐水却会朝着地中海流动。这一来一往之间，地中海的盐分便流入了大西洋。

海水中的盐分会自发地从高浓度的区域向着低浓度的区域扩散，这同样是熵的杰作。如果把大西洋和地中海装到两只杯子里，杯子用管道连通，那么要不了多久，两杯海水的盐度就相同了。它们不会自发变得有序——比如所有的盐分都顺着管道移到同一个杯子中——熵不断增加是自然界的重要规律。

可是生命体似乎有着完全不同的现象。

当婴儿生活在子宫营造的“海水”中时，这种氯化钠溶液就成了他的饮品，流入他的身体。尽管在他身体的任何一个部位都可以找到氯化钠的踪迹，但是浓度并不相同。特别是在细胞的内外，氯化钠的浓度呈现天壤之别：细胞外的浓度要远远高于细胞内的浓度。事实上，正是这种巨大的浓度差让细胞具备了“活着”的功能，可以由此调节血压，也可以传递神经信号。

反之，如果生命体不再“活着”，也将遵循熵增定律，变得越来越混乱。细胞内外的氯化钠不会井然有序，丧失原先的功能。

活着，还是死去，这是一个问题……

在人的一生中，“活着的意义”或许会成为思考最多的问题，很多伟大的哲学思想和文学作品也都因此而诞生。而在物质科学中，“活着”与“死去”之间，最为本质的区别便是熵。衰老乃至死亡，主要是从有序到无序的过程。活着是维持低熵，死去便是任由熵增。

令人好奇的是，生命体何以能够对抗自然界的定律？

## 物质的自组织形式

准确说，人类，还有所有的生命体，至今为止都没能违背“熵增”定律，我们所做的事情，无非是借助于新陈代谢，选择我们需要的食物，靠着食物中蕴含的能量，维持着低熵体系。这项工作我们可以得心应手，就像一条运河，我们可以投入能量维护它，不断地疏浚，不断地修堤，让河流保持畅通，免于熵增带来的冲击。

即便如此，所有的努力也只是起到延缓的作用。所有人都会变老，不能永远保持低熵状态，从而实现“永生”。放眼整个地球上，能够在一定程度上永葆青春的物种，也只是像灯塔水母这样的低等生物，然而它们却也因此牺牲了很多生物功能。

同样的规律，我们还可以在更普遍的现象中发现。一口铁锅和一只电饭煲相比，铁锅的机械结构要简单得多，在同等的质量水平和相同的使用强度下，铁锅会比电饭煲更耐用。换句话说，结构更复杂的电饭煲，保持有序的状态也更难。

因此，人类这般复杂的生命系统，借助于医疗手段的修修补补，能够拥有多达上百年的寿命，这毫无疑问是个奇迹。然而在这奇迹背后，物质自身的规律同样发挥着重要作用。

尽管熵增定律似乎是一个放之宇宙皆准的规则，也就是物质世界会变得越来越无序，但是局部变得越来越有序，倒也还是可能实现的。

想象一杯饱和食盐水溶液，其中只含有水和氯化钠。如果将杯子敞口放置在一只更大的密闭容器中，容器和外界没有任何物质或能量的交换，那么容器内的世界会发生怎样的变化？

容器中的余温会让水分子保持运动，一部分表面的水分子便会挣脱束缚，成为水蒸气，飘散到容器内的每个角落。水分子挥发的这个过程，会带走一部分热量。随后，有些水分子接触到容器的内壁，又会被内壁吸附并凝结成小水滴，同时释放一部分热量。当然，还有一

些水分子会回到食盐水的表面，重新成为溶液的一部分。

这个过程会缓慢地进行，杯中的水逐渐减少，直到达成动态平衡。在动态平衡之下，从杯中食盐水表面脱离的水分子，任何时间都与重回食盐水表面的水分子数量保持相同。相比于最初的状态，杯中水减少，容器其他角落的水增加，整个体系变得愈加混乱，熵自然是上升的。

然而，由于饱和食盐水溶液中水分子的减少，有一些氯化钠也不得不从溶液中脱离，并形成食盐晶体。在海边的晒盐场上，工人们也会利用相似的原理，将浓盐水中的水蒸发，从而获取粗盐结晶。

在这个熵增的密闭容器中，如果我们只观察氯化钠，就会发现，最终形成的晶体比初始溶液状态的氯化钠更有序。换言之，它的熵下降了。

由此可见，即便没有人力的干预，在一个熵不断增加的大体系中，某些物质仍然能够独善其身，甚至有可能变得更加有序。

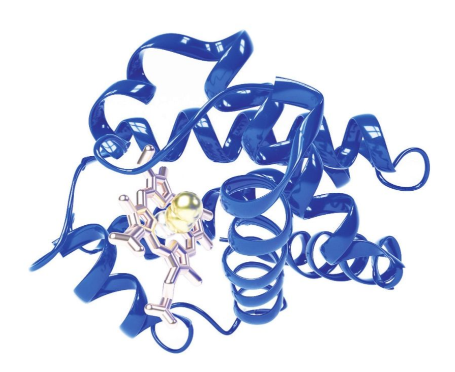
这种现象，源于物质的自组织能力，结晶就是一种普遍存在的自组织过程。

在生物体内，自组织更是一种不可或缺的过程，生老病死都与之相关，蛋白质更是自组织的代表。

当一个个氨基酸连接在一起，形成巨大的蛋白质分子时，一个重要的问题出现了。此时的蛋白质就像是一根麻绳，如果没有任何约束，那么它的熵就会增加，缠绕在一起，变成混乱不堪的绳结。

蛋白质是生命体实现各项功能的重要物质，肌肉、神经组织、免疫系统等等，全都依赖蛋白质分子进行工作。可想而知，胡乱搅成一团的蛋白质分子肯定不堪大用。因此，想要确保蛋白质能够各司其职，就要让特定功能的蛋白质，以相同的方式从长绳编织成绳结。比如人体血液中充斥的血红蛋白，主要功能是携带氧气，只有确保每个血红蛋白的结构全都一样，才能保证它们拥有相同的供氧能力。如若不然，就有可能引发严重的疾病。事实上，有一种疾病被称为镰刀型细胞贫血症，即原本很像柿饼的红细胞扭曲成镰刀状，供氧能力显著

下降，其背后的原因，就是血红蛋白的结构发生了异化。



血红蛋白分子的模型

生命体中的蛋白质想要形成特定结构，依赖的就是蛋白质的自组织特性。不同于一般的分子，蛋白质的长链并不会完全自由地扭曲，而是在氢键的作用力之下，局部自发形成一些特定的形态。它们有时会像弹簧那样卷曲，有时又会像被子那样折叠起来，为了方便表述，研究者给它们起了诸如α-螺旋或β-折叠之类的称呼，这样我们就很容易想象这些蛋白质大致的模样。

不过，相比于共价键，氢键的力量仍然有限，仅仅依靠不能保证蛋白质每一次组织都能准确进行，更不能保证自组织的结构能够稳定。对生命体而言，这是一种不可控的隐患，因为熵增定律的存在，任何不稳定的要素都可能会导致有序的结构变得混乱。如今，我们已经知晓，阿兹海默症的发病缘由，就是某些蛋白质在折叠时发生了错误，变得有些无序。阿兹海默症俗称老年痴呆症，年龄高低与发病率紧密相关，由此我们可以亲眼见证，熵会随着时间推移而增加。

通常情况下，胱氨酸会作为蛋白质结构中的奇兵，如同盆景艺术中的支架那样，让蛋白质的结构变得更加稳定。胱氨酸实际上是由两个半胱氨酸构成，每个半胱氨酸中都含有一个被称为“巯基”的结构，也就是一个硫原子与一个氢原子形成的基团。巯基就如同是半胱氨酸的手掌，在一定的条件下，两个巯基可以紧紧地握在一起，让蛋白质彻底定型。

蛋白质自发定型的过程就如同是“自来卷”的头发，生长出来便是

特定形态的卷发。事实上，动物的毛发也是一种蛋白质，它们能够按照特定规律卷曲，内在原因也正是蛋白质的自组织现象。不少人因为爱美而烫发，无论是从直发烫成卷发，还是从卷发烫成直发，在分子层面上看，都是对蛋白质的结构重新定型——打开由巯基构成的绳扣，再换一个方式重新系上，毛发的形态就发生了变化。

人体中有数万种不同结构的蛋白质分子，每一种蛋白质各司其职，都要依赖它们的自组织能力。进一步说，蛋白质的自组织现象，也只是生命体中众多自组织现象的缩影。骨骼的主体是羟磷灰石，但它还需要填入其他一些物质才能确保自身的强度与韧性，没有强大的自组织能力，这种复杂的结构无法形成；植物的树干中布满了微管，它们并行不悖地给树叶输送养分，这种有序的结构需要自组织能力；所有生命体都必须将自己的特征遗传给下一代，为了实现这一目标，需要的还是自组织能力。

正因为物质具备自组织特性，生命才能在熵增的趋势中保持相对有序，人体这样复杂而精密的结构，可以连续运转上百年的时间。即便旧的生命体停下了脚步，也还会有新的生命体延续下来。

然而，在生命与生命之间遗传，除了物质本身的功能外，还需要“信息”的传递。

## 遗传密码

人体有46条染色体，它们两两配对，分布在血红细胞以外的几乎所有细胞之中。这23对染色体，就如同是一个人的生命档案——这个人如何成长，如何发育，在很大程度上都需要翻阅这座微型的档案馆，参考了其中的数据之后才能进行。

19世纪末，科学家们已经能够对细胞进行细致的研究，通过一些方法区分出细胞内的不同组分。当他们使用一些染料对细胞进行染色时，发现有一些成分很容易被染上颜色，故而称之为染色体。

进一步的研究发现，染色体似乎和细胞的繁殖存在直接关联。以一般条件下的人体细胞为例，23对染色体被精巧地布置在一片狭小的

空间里。这块独立的小空间被称为细胞核，核外有一层薄膜，将染色体与其他物质分隔开来。不过，此时的染色体并不能被染色，人们也更习惯于称它们为染色质。

人的一生中，细胞的数量总是在不断地增加。所有人的起点都只是一个受精卵，那不过是一个细胞而已，然而成年人的身体却是由数十万亿个细胞构成，只有细胞不断地分裂，才能发展出如此庞大的细胞集合体。即便是已经停止发育的成年人，新陈代谢依然会产生一些新细胞，同时淘汰一些已经凋亡的细胞。

由一个细胞分裂出更多的细胞，并不只是简单的数量变化，更重要的是，新生的细胞要和旧有的细胞之间保持一致性。换句话说，自然条件下，每个人都只能生长出属于自己的细胞。因此，细胞通常需要进行复制的过程——通过分裂得到的新细胞应与旧细胞看上去一模一样。这个问题说起来容易，可细胞毕竟不是一张身份证，放到复印机下就能轻松制作出副本。对细胞而言，所谓的复制是要将细胞内所有的物质都妥当分配。事实上，细胞分裂的过程，有点像是封建时代的两兄弟分家，家长需要公平地把各种财产一分为二。

细胞之中，唯一难以分配的“财产”便是细胞核，更确切地说，是细胞核内的染色质。为了避免“发生矛盾”，当细胞进入分裂的周期时，细胞核内就会预先进行复制，在很短的时间里，23对染色质摇身变成46对，每一对都可以找到和它完全相同的副本。

此时，染色质已经可以被染色，变成真正的染色体，并且通过它们特有的自组织方式，形成了粗壮的棒状结构——这些变化都让研究者更容易观察到细胞内正在发生的事情。当细胞完成分裂后，新生成的两个细胞，各自获得了一套染色体副本，它们与原先细胞的染色体完全一致。

这并不是细胞增殖的唯一方式。当新的生命诞生之时，胚胎细胞如何出现，更加引人注意。

对人类来说，胚胎细胞由两个细胞合并而成——一个是来自于母体的卵子，另一个则是来自于父体的精子，它们结合在一起，成为受精卵。

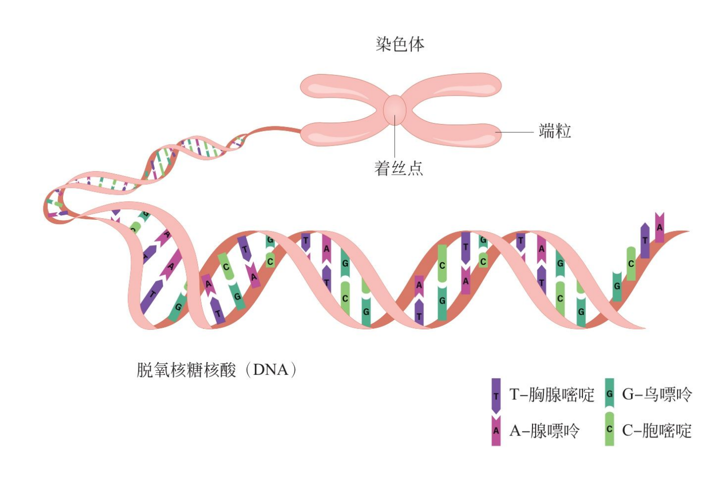
在受精卵的形成过程中，染色体的变化方式不同于常规。卵子和精子中都只有23条染色体，也就是正常细胞中的一半数量，每个染色体都只是原先一对中的其中一个。而当卵子和精子结合时，它们又将重新组成23对染色体。

正是这样的分配方式，新生命的很多属性都能够确立下来，例如性别。在这23对染色体中，有一对染色体被称为性染色体。不同于其他被常染色体的22对染色体，性染色体有X型和Y型两种，体积差别很大。女性的两条性染色体都是X型，而男性则分别是X型和Y型。因此，卵子中的性染色体必定是X型，精子却有X型和Y型两种类型。最终，与卵子结合的精子中所包含的性染色体类型，会决定新生儿的性别。

在观察了染色体数十年后，科学家们越来越在意它们对于遗传的重要性。美国实验坯胎学家、遗传学家托马斯·亨特·摩尔根（Thomas Hunt Morgan，1866年—1945年）通过对果蝇染色体的实验遗传研究，发现伴性遗传规律，最终确定染色体就是遗传信息的载体，并因此获得1933年的诺贝尔生理学或医学奖。

在此之前，人们早就知道，染色体主要是由核酸和蛋白质两大类物质组成。更进一步的研究证明，承载遗传信息的物质就是核酸。

我们已经知道，核酸是生物大分子的一类，由磷酸根连接而成的长链分子，每一个磷酸根都可以连接一个核苷酸分子。核苷酸分子中有两个部分：一部分是糖分子，通常是核糖或脱氧核糖；还有一部分是碱基分子，也就是嘌呤或嘧啶分子。人体储存遗传信息的核苷酸所含的是脱氧核糖，最终形成的核酸分子，便被称为脱氧核糖核酸（DNA）。与之相对，以核糖作为构成单元的核酸分子被称为核糖核酸（RNA），也有一些生命体以它作为遗传信息的载体。




染色体与DNA

蛋白质通常都是很大的分子，但是相比于DNA，它们还是小得出奇。人体中的46条染色体，每一条都包含一对核酸分子。这些DNA分子的长度不一，其中最长的分子，其骨架竟然是由大约2.5亿个磷酸根连接而成——每一个磷酸根还都连接了一个带有碱基的核苷酸。

正是这些碱基，成为生命体的遗传密码。

和绝大多数生命体一样，人类的DNA分子含有四种碱基，它们分别是腺嘌呤、鸟嘌呤、胞嘧啶和胸腺嘧啶，通常被简写为A、G、C、T这四个字母。每个人都有属于自己的碱基顺序，不同的碱基顺着DNA分子排列，就构成了一段密码，例如ATAGCCA……这样的密码足有上亿个字母，其中包含的信息量也非常惊人。

并不是所有的密码都有意义。实际上，只有少部分密码可以被翻译出来，这样的密码被称为基因。基因包含的碱基密码数量并不固定，通常包含数千个碱基。每三个碱基，就可以被翻译成一种氨基酸，这样一来，一个基因就可以决定一段氨基酸的序列，让它们连接成固定结构的蛋白质。人类的基因数量有两万多个，但是人体内蛋白质的种类却远超这个数字，它们到底是怎样实现的，现在依然很令人好奇。

我们现在已经知道的是，为了保存好这些基因，生命体做出了哪些努力。

1952年，英国化学家和DNA研究先驱罗莎琳德·富兰克林（Rosalind Franklin，1920年—1958年）拍摄了一张著名的DNA的X射线衍射照片，她称之为“51号照片”，照片的主角正是DNA分子，呈十字状排列的暗纹为DNA分子的螺旋状结构提供第一个证据。第二年，两位英国科学家詹姆斯·杜威沃森（James Dewey Watson，1928年—）和弗朗西斯·哈利·康普顿·克里克（Francis Harry Compton Crick，1916年—2004年）对照片进行解读，终于参透了DNA的奥秘，合作提出了DNA分子的双螺旋结构模型，由此开辟了分子生物学的新时代。事实上，每条染色体中的一对DNA分子，是以双螺旋结构结合在一起——就像是扭曲的梯子一样。在这个梯子中，两条侧边分别是两个DNA分子的磷酸根骨架，横档则是它们连接的核苷酸。核苷酸上的碱基会因为氢键两两相吸，A总是和T结合，而G总是和C结合。这样一来，这两个DNA分子就会完全互补，即便一个DNA分子完全消失，借助于另一个DNA的密码，也很容易将失去的那一个重新构建。有了这个机制，DNA分子将遗传信息完好地保存，很难出现错误，生命体也因此得以将自己的基因传递给下一代，繁衍生息。

遗传物质能够传递生命的信息，这一点本身并不让我们感到意外。我们可以把DNA想象成是一张纸条，上面密密麻麻地写着AGCT，生命通过自己的解码系统，利用这张纸条施工，最后搭建出不同的生命形式，或许是人，又或许是一只猫。但是，物质能够在生命的层面上对信息进行记录和解析，似乎启发了我们再次思考那个一开始就被提出的问题：物质是什么？

于是，整个宇宙中的这些物质，就这样搭建出我们已知最复杂也最精细的结构——或许称之为“系统”还要更合适一些。这些生命系统不停地运转，给物质世界带来了蓬勃生机，却也带来了新的危机。

然而，物质以及物质演化而成的生命形式，它们是否能够“可持续”地存在下去，这是一个更值得思考的问题。

# 8 冲突与重生——物质世界会终结吗

## 谁杀死了恐龙？

霸王龙在它的“园子”里徜徉，它那十余米的体型，配上尖锐的牙齿与厚实的皮肤，直叫周围的生物畏而远之，没有谁敢打它的主意。

地球诞生以后，包括火山、地震在内的各种地质事件层出不穷，阳光携来的能量按动了风云变幻的开关，月升月落悄悄地诱动海水引发潮汐，就连太阳系中流浪的小行星也不时地修理一下地球。斗转星移，地球表面的恶劣环境总算有了改善，繁荣增长的生命又开始书写着属于自己的历史。于是，地球的地质环境在这些力量的塑造之下，不断地改换模样，一起变化的还有地壳中的物质种类。

为了能够更好地研究地球演化的过程，地质学家用一套特殊的纪年方式，为地球编制了“地质年代”谱。

这个“地质年代”谱是描述地球历史的时间表，科学家根据其等级系列，确定所有地层的年代标准尺度。在这个纪年法中，地球经历的45亿年沧桑，被划分成三个“宙”，分别是太古宙、元古宙和我们正在度过的显生宙。显生宙起始于大约5.7亿年前，它又被分为三个“代”：古生代延续到大约2.5亿年前，中生代接棒到6 500万年前，新生代又把故事续写至今。每一个代被划分成不同的“纪”，每个纪又是由好几个“世”构成。

温暖的中生代，铸就了生命繁衍的奇迹，值得一提的是，在三叠纪出现了哺乳动物。无论是三叠纪、侏罗纪还是白垩纪，都是爬行动物的时代，它们留下了很多大型动物的化石，尤其是属于爬行动物的恐龙，在中生代尤其繁盛。

而在人类创作的文艺作品中，恐龙几乎成了侏罗纪的生命代言人，霸王龙就是那个时代的霸主——然而，对恐龙家族的研究结果却

证明，霸王龙直到白垩纪才出现。这是披羽毛、长翅膀的恐龙或者说早期鸟类兴起的时代。

在霸王龙的统治之下，还有很多恐龙物种漫步在陆地上的每一个角落，它们同样是中生代不可或缺的成员：背挎长矛的棘龙挑战霸主的地位，长有头角的三角龙不甘沦为对手的食物，近十层楼高的梁龙伸长脖子傲视群雄，长有翅膀的翼龙正在训练着滑翔技术。

然而，在一场灾难过后，它们全都消失了。所有的恩怨，也随之灰飞烟灭，大概只有粗糙的飞行技术，成为鸟类后来统治天空的法宝。



根据化石想象的三角龙与霸王龙对决

恐龙灭绝的原因，至今众说纷纭，其中最流行的说法，是一颗直径大约10千米的小行星撞击了地球，由此引发了一连串灾难性的后果，包括恐龙在内的很多生物都在此过程中灭绝了。

如今，我们还可以在地球上找到这颗小行星的遗骸，它们也成为现代人研究生命演化的重要依据。这并不奇怪，正如我们已经知道的，物质不会凭空消失，只会转化成新的形式保留下来。因此，哪怕只是一颗小小的流星，也可以在地球上留下痕迹。这些流星通常会在进入地球大气后，因为空气的摩擦而剧烈升温，最终发生解体或燃烧，最后只剩下细小的颗粒，飘散在空气中，最后落到地面。稍大些的流星，或许还能以陨石的形式落在地面，有的陨石以岩石为主，有的陨石则以金属铁为主。在人类还未学会冶铁技术以前，就是这些陨铁，提供了当时最顶尖的金属材料以供打造器物。

总之，大气层对地球而言，就如同是一张巨大的防护网，它迫使各类随意闯入的不速之客减低速度，甚至直接将它们“没收”，避免它们对地球造成巨大的破坏。

但是，6 500万年前的这颗小行星实在是太大了，大气层也无可奈何，只能任由它横冲直撞。事实上，不只是空气组成的大气层，就连岩石组成的地壳，也没能阻挡住这颗庞然大物——它落地的一瞬间，就足以覆盖一座现代的大城市。而它的惯性又是如此巨大，顺势就砸入了地幔之中，只在如今的墨西哥，留下一处直径大约180千米的陨石坑，这也成为科学家们还原当时那场灾难的依据。在地球上，还没有哪次地震或火山会有这样恐怖的威力，也就难怪它居然能够终结这漫长的中生代。

对于恐龙生活的那个世界而言，并没有什么建筑物会受到损失，但是生命繁衍所遭受的危机，却真实地降临了。

巨大的撞击，迅速改变了地球的地质环境，地震和火山接踵踏来，海面更是漾出了有如山高的巨浪。剧烈碰撞带来的高温，还引发了长久不灭的山火，整个地球都充满了末日焚烧的气息，灰烬夹杂着小行星撞击时扬起的尘埃，随风扩散到各个角落，又降落到了地面。

霸王龙钟爱的园子被毁了。

和1815年印尼那场史无前例的火山爆发相仿，被尘雾笼罩的地球，已经不能接收到充足的阳光，大地陷入荒芜的冬天，光合作用显著减弱，大批植物因此不再生长繁殖，甚至就此灭绝。

对于那些食草的恐龙而言，植物不再生长，意味着食物开始短缺。它们为了生存艰难跋涉，但是地球虽大，却已没有任何一处能够给它们提供水美草肥的栖息地。终于，最后一头饥肠辘辘的食草恐龙也倒下了。

这个消息，对于霸王龙这样的食肉恐龙而言，如同是宣告了灭绝的倒计时。靠着随处可见的腐肉，它们或许还能挨过一段时间。可是，新的猎物不会再出现了，它们也支撑不了太久。

总之，恐龙时代结束了。

当我们总是反复讲述这段故事的时候，可曾想过，杀死恐龙的，真的就是那颗出乎意料的小行星吗？

的确，它把地球砸出了一片让人触目惊心的伤疤，无数的生灵因它流离失所，其中有很多都走向末路，从此只能蜷缩在土块岩石里，历经千万年的变化，也许从此深埋海底，也许有幸重见天日，成为博物馆中的“化石”。

恐龙的命运是悲惨的，所有恐龙家族都没能幸免于难。但是，恐龙又是幸运的——和恐龙一同灭绝的物种，都没有像它们这样受到关注。

每当我们郑重其事地仰望恐龙的骨骼化石，都会发出一阵惊叹：它们的体型，实在是太大了。

小行星没有光临的时候，硕大的恐龙是这个世界上最主要的消耗者，从微生物到植物，再到各种食草动物，它们勤勤恳恳地创造出各种食物，最终都献祭给那些凶猛的恐龙。显然，这就是一个围着恐龙组建起来的生物圈系统，物质在系统中流动，驱动生命繁衍不息。

而当灾难来临之时，这套系统渐渐失灵，恐龙赖以生存的园子也随之土崩瓦解。

恐龙消耗了最多的资源，自然也就成为这场灾难中最艰难的群体。由俭入奢易，由奢入俭难，这道理不只是在《红楼梦》中贾府的“大观园”里成立，同样也适用于恐龙生活的那个侏罗纪公园。

相反，发迹于中生代的哺乳动物，却在这场灾难中因祸得福。它们大体和现在的老鼠差不多，弱小的身躯根本不是恐龙的对手。只不过，恐龙属于爬行动物，体温会随着外界变化，夜间便休息了，体温能够保持恒定的哺乳动物就利用这个间隙寻找一点食物残渣。

就是因为生存需求如此卑微，哺乳动物才坚强地挺过小行星撞击后的黯淡冬日。进入新生代后，恐龙消失空出来的生态位，便由哺乳动物填补。如今，作为哺乳动物中的佼佼者，人类认为自己是这个星球的主人。

与其说，恐龙亡于那颗小行星，还不如说是倒在了自己的大体型

上。地球虽然是一颗直径超过1万千米的大行星，物质资源异常丰富，但它需要妥善地经营，无法供恐龙这样的生物持续挥霍。相比于地球，10千米的小行星就如同在我们脚边爬行的蚂蚁，但它却足以扰乱整个地球的物质输送系统，诚如一只蚂蚁也可以让人感到疼痛一般。

然而，即便没有这颗小行星，恐龙那粗放的生存方式，也不可能长期持续。地球表面的任何变化，都可能会让它们遭受灭顶之灾。最终，它们即便不落得黯然退场，大概也只能像鲸那样回到生命最初发源的海洋中，那里暂时还能给这些大体型的生命提供足够多的物质。

但，这也只能说是暂时。

## 生物链的物质流动

北冥有鱼，其名为鲲。鲲之大，不知其几千里也。

在庄子的《逍遥游》中，有一种纵横数千里的大鱼，生活在最遥远的北方大海之中。鲲还会变化成一种叫做“鹏”的大鸟，挥舞着数千里宽的翅膀，朝着南方的天池飞翔。

当然，现实中不可能存在这样的大鱼。迄今为止，人类在海洋中找到的现存鱼类中，最大的要数鲸鲨，它的身长可以达到10余米。而在海洋之中，我们还能找到蓝鲸这样的哺乳生物，它们的体长可以超过30米，体重更是可以达到180吨，相当于三节复兴号列车的重量。实际上，在已知的研究结果中，蓝鲸也是地球诞生至今存在过的最大生物体。

地球上是否还会演化出更大的生物体呢？这个问题，并不只会吸引庄子这样的哲学家，同样也让科学家们百思不得其解。

可以肯定的是，如果还有这样的生物，那它一定不会出现在天空中，自然也就不会是“鲲”化作的“鹏”。

相对而言，地球表面的空气还是太过稀薄，即便是在海平面附近，其密度也不足水的七百分之一，若是爬上高山，空气就更稀薄

了。正因为此，空气不可能提供太大的浮力，生物体若是要天际遨游，就需要耗费更多的力气，舞动翅膀，这样才能保持空中的姿态。

正因为此，如果鸟类的体型过大，例如鸵鸟和渡渡鸟，就只会在地面生活。即便是丹顶鹤这样能够飞起来的大鸟，通常也需要像跑道上的飞机一样，在一阵助跑后才能离开地面。

不只是天空，陆地上也承载不起特别庞大的身躯。因为万有引力的存在，任何动物都需要支撑起自己的重量。这对于昆虫大小的动物来说，的确不算什么，但是对于人类这般体型的动物而言，自重的影响就无法忽视了。正如我们已经知道的，坚硬的羟磷灰石组建了我们的骨架，让我们能够做出奔跑、跳跃这样的动作。但是，当一个人过于高大或肥胖时，即便是这样一副骨骼，也会变得似乎有些脆弱。

目前，陆地上最大的生物是象，它们的体重可以超过6吨。相比于已经灭绝的恐龙，大象只是算是小个子，而最重的恐龙大约有100吨，这也是已知的陆地生物极限体重。

未来的地球上是否还会出现更大型的陆地生命？科学家们对此问题的答案并不看好。这不仅是因为庞大的身躯对于骨骼带来的强大压力，更是因为，恐龙的灭绝，实质上也宣告了大体型生物对抗风险的乏力。

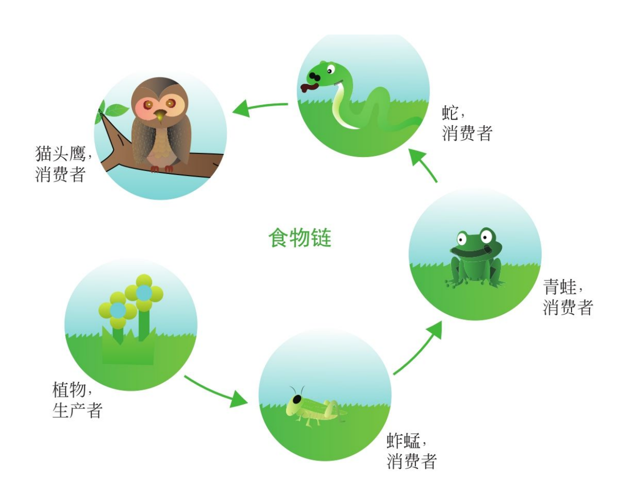
寻找更大生物，在地球上就只剩海洋这样的场景了。

海水可以提供更大的浮力，包括人类，都可以非常轻松地浮潜在海水之中，骨骼需要承受的压力也更小。事实上，很多人在骨骼受伤后，也会通过水下行走来帮助恢复，正是借助水的浮力，从而抵消了自重对骨骼的巨大压力。反之，若是鲸搁浅了，自身的体重就有可能会导致骨折。

不过，海洋生命更容易演化出“大块头”，更重要的原因，还在于充足的食物供应，这是地球物质流动规律所决定的。

在陆地上，无论是大象还是恐龙，最大的动物都是以植物为生。对它们而言，只要能够找到自己喜欢吃的树叶或青草，就可以安逸地度过一天又一天。植物将二氧化碳和水转化为葡萄糖以及其他有机

物，食草动物都以此为生，食肉动物又以食草动物为食。整个过程形成了食物链，物质就在这个食物链中发生了转移。




食物链并不总是这样简单，它还可以形成更多层级。一条经典的食物链发生在田野之间：白菜作为初始的生产者，利用光合作用制造出有机物；蚂蚱在田间跳动，找到适合自己的“自助餐”座位吃个饱；青蛙早就蹲守在低处，舞动的蚂蚱恰好给了它捕食的机会；蛇是青蛙的天敌，它没有错过美餐的时间；这一切都被夜间巡视的猫头鹰看在眼里，它一个俯冲就将蛇叼了起来。

在这条食物链中，白菜、蚂蚱、青蛙、蛇、鹰，物质在它们之间依次流动。蚂蚱可以消化白菜中的纤维素，转化自己所需的各种有机物，特别是蛋白质，而青蛙、蛇和鹰都能够消化其他动物的蛋白质。不过，物质的这种流动效率并不是很高，每一级食物链自下而上时，只有大约1/10的物质或能量会被利用。换言之，一只蚂蚱吃下10片叶子，也只有一片叶子转化为它所需要的成分，其余的物质，只是为了让它能够完成吃叶子的动作。

于是，随着食物链向上传递，物种所含的物质总量也会显著减少。白菜远比啃食它的蚂蚱更多，它也必须做到这一点才能够继续繁殖。同样，食物链底端的动物往往会形成数量优势，确保自己不会在生存竞争中灭绝。当蚂蚱形成蝗灾时，一群蚂蚱就有数亿只之多，远非青蛙的数量可比。至于这条食物链的最上层，鹰的数量就要少得

多，难得让人发现它们的身影。

陆地生物想要拥有更大的体型，首要的前提，是食物必须十分充沛，这样才能有足够多的物质，转化为这些动物身体的一部分。能够满足这样要求的食物，就只有遍地生长的野草了。所以，我们也就不难理解，为什么只有食草动物才能生长出更大的体型了——它们的食物更充沛，它们在食物链所处的位置更低，可以被它们利用的物质自然也就更多。

和陆地相比，海洋真是个大宝库。正所谓“海纳百川”，很多化学元素在地球上的循环，都是以大海为最终的归宿，特别是氯和钠，它们从岩石溶解到雨水中，顺着江河入海，让海水变得越来越咸。而像碳、氮、磷这样一些生命中的主要元素，同样也参与了这样的循环，融入海洋。

正因为此，世界上几乎每一条大河的入海口，都会成为海洋生物的宴会场。在中国，长江入海口的南北两侧，分别形成了舟山渔场和吕四渔场这样的重量级渔场，就是因为长江经过6 000多千米的奔腾，带入海洋的，不只是泥沙，更有海洋生物赖以生存的各种元素。

不只是拥有更多的物质，海洋相对简单的环境，也让大型动物更有机会像饕餮那样不停地进食。就像蓝鲸，它只需要张开大嘴、滤食海水，海水中的巨量磷虾，就被它吞入腹内。有了如此充沛的食物来源和高效的进食方式，才成就了蓝鲸这样的深海巨兽。

饶是如此，相比于抹香鲸、虎鲸这样的齿鲸，仅以小型动物为食的蓝鲸，在生物链上的位置还是要更低一些。这也说明，它获取海洋物质的效率也要更高。

长期以来，人类一直都在海洋中寻找比蓝鲸更大的生物，但都是无功而返。借助于各种深海探测器，我们有机会到神秘的海底一探究竟，去寻找最终的答案。但是从物质流动的角度而言，我们无法想象，在海洋某个无人所知的角落，会存在一种“鲲”，它可以几乎无限地获取自身发育所需的物质，生长出令人惊叹的体型。

如果有，那可能就是我们人类自己。

## 被浪费的物质

在生物界中，人是一种大型动物，体型超过人类的生命，除了各种树木以外，能够叫出名字的现存动物寥寥可数。

在生命物质的大循环中，植物通常是生产者，而动物则是消费者。人类也是如此，我们需要吸入植物产出的氧气，消化它们合成的淀粉。当然，人类并不只以植物为食，相比之下，对肉食的兴趣还要更浓厚一些。到了现在，我们似乎已经走到了食物链顶端。

正如我们已知的，在食物链处于较高地位的动物，体型越大，自然也就需要更多的物质来支撑它。要紧的是，地球上活着的人类数量如今已经超过80亿，这个数字远远超过其他大型动物。以人类的近亲黑猩猩为例，它的体型和人类差不多，但它们的数量却只有几十万只而已。

所以，仅仅是为了像其他动物那样在地球上生存繁衍，人类就会消耗最多的资源，产出最多的废物。

不仅如此，人类在进入工业社会以后，对物质生活的追求也越来越高，对物质的利用效率却在不断降低——我们正在浪费越来越多的资源。

人类的祖先经历过茹毛饮血的年代，那时他们吃白菜的方式，和蚂蚱并没有太大的差别。在这个时期，白菜生产的有机物，可以被人类有效利用。

后来，火被人类掌握，我们的祖先又学会了烹饪。对食物进行加热，可以让其中的一些成分发生转化，特别是难以水解的蛋白质。如此一来，食物在被消化之后，会有更多的物质转化为身体所需。因此，有了火以后，物质从食物转化到人体内的效率，又上升了一个台阶。

然而，这样的效率提升并没有延续下去。

当李比希解开了植物生长的元素之谜后，肥料成了植物培育必不可少的一类物质，特别是最紧俏的N、P、K肥料。

人类在找寻NPK肥料时发现，氮、磷、钾这三种元素，植物虽然都很容易缺乏，但是缺乏的原因却各不相同。

氮是空气中最丰富的元素，只是氮气的化学性质太稳定了，除了豆类以外的绝大部分植物，都无法利用空气中的氮气。因此，只要开发出一种办法，把氮气转化为其他含有氮元素的物质，就可以源源不断地生产出氮肥。这项工作后来由李比希的同胞弗里茨·哈伯（Fritz Haber）解决了，他提出的合成氨工艺，可以让氮气和氢气发生反应，转化为氨气，而氨气就可以被加工成各类氮肥，能够被植物快速利用。

相比之下，钾元素倒是很容易被植物吸收，而且它也很常见——在地壳中，钾的丰度可以排到第七位，略低于海水中的“霸主”钠元素。然而，钾元素在地球上的分布并不均匀，有些地区过于丰富，有些地区却又过于寒酸。更重要的是，那些钾含量丰富的地区，常会因为气候因素不适合耕种，例如中国的罗布泊地区，就因为气候干旱成了荒漠戈壁。所以，开发钾肥的办法，就是给它们搬家，从罗布泊这样的地方送到农田。如今，曾经荒无人烟的罗布泊已经建立起雄伟的钾肥生产工厂。

最让人们难以释怀的元素是磷。在地球上，磷元素的含量算不得太多，而且分布也极不均匀。实际上，大多数生物在自然条件下生长时，都会面临缺磷的窘境。植物如此，动物如此，就连海水中微末的浮游生物也是如此。

在太平洋上，有一个名叫瑙鲁的国家，是世界上面积最小的岛国，比中国最小的省级行政区澳门还小了几分。

别看瑙鲁的地盘不大，却拥有异常丰富的资源。它刚好位于很多候鸟迁徙的路线上。茫茫无际的太平洋上，这座袖珍岛屿成了这些鸟歇脚的中转站，在这里进行补给后，再赶往下一个目标。

因为这个岛实在太小，以至于鸟类在此处栖息时，只能非常拥挤地聚在一起。如此一来，这么多鸟排泄的粪便，也在岛上集聚起来。日复一日，年复一年，巨量的鸟粪就堆得如山一般，成为岛上岩石的组成部分了。谁曾想，这些鸟粪，居然是一种宝藏肥料。

当瑙鲁的鸟粪化石被发现后，人们惊人地发现，它们其实已经转化成优质的磷酸盐，可以称得上是天然的磷肥。

就这样，围绕着这些鸟粪，德国、英国、日本、澳大利亚数次争夺对瑙鲁的控制权，这样就能抢占这些磷肥资源。直到1968年，在联合国的支持下，瑙鲁才获得自由，成为主权独立的国家。

可是，在近一百年的磷肥开采权的争夺中，瑙鲁的磷资源早已被开采殆尽。瑙鲁人在经历了开采磷肥带来的短暂巨富后，很快陷入贫穷，如今更是成为犯罪分子的天堂。

问题是，那些被开采出来的磷元素都去哪儿了呢？

这是一个令我们感到恐慌的故事。

人类在施用肥料的时候，只有极少部分会被植物利用。特别是磷肥，因为不当滥用和吸收效率的局限，超过95%的磷元素都不能转化为植物的一部分。

换言之，如果一块鸟粪中含有1 000个磷原子，我们经过开采，大约会有800个磷原子能够被作为磷肥使用——这样的采集效率已经不低了。这些磷肥被运到白菜地后，只有不到40个磷原子会被白菜吸收，而当这些白菜成为人类的食物时，我们只利用了其中大约4个磷原子。如果我们没有直接吃白菜，而是把白菜给猪吃了，等我们再吃猪肉时，那么能够被我们利用的磷原子甚至还不足1个。

那些被浪费的磷原子，就会在江河的作用下，被排放到海水中。此时，早就饥渴难耐的浮游生物终于盼来了它们日思夜想的磷元素，报复性地开始繁殖，甚至能够在海边引发一场赤潮。那赤红的潮水，似是混入了鲜血，让人不寒而栗。



赤潮

事实上，赤潮对很多生命而言都是死亡的讯息，海水中溶解的氧气会在短时间内被消耗殆尽，大量鱼类、虾蟹、贝类、爬行动物乃至哺乳动物都会因为缺氧而丧命，酿就一场生态灾难。

一阵喧嚣之后，逝去的生命体裹挟着大量的磷元素沉入海底。

现在，你应该知道，为什么尽管海纳百川，海洋对巨型生物而言，也不过只是暂时安全的栖息地了吧？

## 天地间，人为贵

最近半个世纪来，人类已经逐渐认识到，如果按照目前这样粗放的物质利用方式，我们怕是也要重蹈恐龙的覆辙——我们已经成为地球物质流动过程中最显著的环节之一，远甚于侏罗纪时代的恐龙，以至于很多科学家都认为，按照地质年代划分，现在应当属于显生宙新生代第四纪人类世，人类活动已经让地球表面发生了巨大变化，或许不亚于6 500万年前那场小行星撞击地球的事件。

为了争夺资源，人类曾经爆发过无数次争端：有一些是和其他物种竞争，剑齿虎和猛犸象这些生物再一次因为体型过大而永远消失；更多的冲突来自人类内部，智人这个物种融合了尼安德特人，又在一次次“内战”中衰退、重生。

实际上，世界人口并非是一路逐步增长到现在的数十亿，而是总在曲折发展，有时候甚至还走到过灭绝边缘。距离我们最近的一次危机发生在大约11万年前，地表温度骤降，步入“大冰河时期”，直至大

约1.2万年前才结束。

在一片白茫茫的冰冻星球上，任何物质都变得极为珍贵。这场变故，正是剑齿虎和猛犸象灭绝的直接原因，但是人类并没有因此获得足够多的战利品，最艰难的时候，整个地球上也只剩寥寥数万人。

即便是在大冰河时期结束后，也还是会时不时地进入小冰河时期，平均气温比正常情况下低了一两度。这样的温度变化看似不大，却会让粮食作物严重减产。历史学家在对明朝进行研究时，猜测明朝末期中国人口骤减，除了战乱的原因外，恐怕主要还得归结为当时正处于小冰河时期。

如今，世界人口如此庞大，而我们对资源的消耗与浪费又是如此严重，一旦地球表面的物质流动过程出现差池，必然就会带来更为严重的冲突——又有谁知道，在这样的冲突之后，是否还有重生的机会呢？

所以，人类开始学会自救，也必须要开始自救。

像磷元素这样的物质，并不只是瑙鲁的鸟粪出现了枯竭，全世界的磷矿都已告急。人们之所以恐慌，是因为按照生命的规律，磷元素不可或缺也无可替代，但是对于那些沉入海底的磷元素，我们却又只能望洋兴叹。

于是，新的肥料技术正在发展，以便植物能够更高效地吸收这些磷元素。更重要的是，通过环境治理，可以让磷元素的流失变得更慢一些。还有一些方法，是从我们的废弃物中回收磷元素，其中也包括人类的排泄物——某种程度上说，这也是鸟粪带给人类的启发。

如今，地球上探明的磷资源还够人类使用大约50年，但是，只要我们提高效率，就可以让这个时间显著延长——如果1 000个磷原子中，我们能够利用的部分从4个提升到40个，那就意味着，留给我们的时间还有500年。

但是，还有很多物质，它对我们的意义并不只是简单的资源。

1973年，第一次石油危机爆发，人们突然发现，原来石油也面临枯竭的风险。实际上，石油是深埋于地下数百万年乃至上亿年的生物

化石，它们曾经可能是藻类、细菌，也不排除会是一些动物，其中也包括恐龙。这些生物体中富含的脂肪类物质，会因为深埋于地下难见天日，最终转化为以碳和氢两种元素为主的有机物，这便是石油。

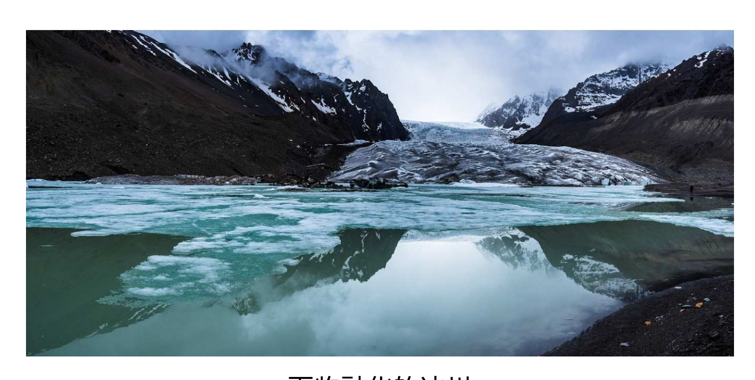
所以，如果我们持续不断地开采石油，那么总有一天，石油会被我们消耗干净。因此，当学术界揭示这个奥秘后，人们立刻意识到问题的严重性。在现代工业体系下，石油可以用来生产衣服面料、汽车轮胎和各种塑料，更是交通工具中的主要燃料。没有了石油，生活会变得难以想象。

几十年过去了，石油依然是人类社会中的战略资源，但人们也惊奇地发现，总是会有新的油田会被发现，石油似乎源源不断。另一方面，关于石油的成因，也出现了更多的理论依据，似乎并不总是需要花上数百万年才能形成。如此看来，石油枯竭的风险，还不足以形成真正意义的“石油危机”。

然而，另一个幽灵却在撬动着人类敏感的神经。

当我们尽情燃烧了上百年的石油与煤炭后，地球却在悄然发生着变化。2022年，大气层中二氧化碳的平均浓度将会突破0.42‰，相当于空气中每100万个分子中，就有420个是二氧化碳分子。这个数字，大约是工业革命发生前的1.5倍。事实上，仅仅是在7年前的2015年，这个数字也才刚刚突破了400而已。

毫无疑问，我们正在加速向空气排放二氧化碳，这已经远远超出植物通过光合作用所能吸收的程度。



面临融化的冰川

二氧化碳是一种温室气体，它让地球更容易吸收阳光中的能量。因此，越来越多的二氧化碳，就好比是把地球裹成了蔬菜大棚，地球表面的温度会因此显著增加。

当地球表面的温度突然下降时，小冰河时期会让粮食减产。但是，当地球温度突然上升时，由此带来的全球气候变暖也会造成同样的结果。比这更令人担忧的是，更温暖的空气会让气候变得更加极端，冰川融化更是会导致海平面上升。到那时，瑙鲁面临的问题已不再是消失的鸟粪，而是整个岛屿都可能会消失。

而这一切，都源于我们对资源的过度消耗。

追求更好的生活，并不能被称为贪婪。但是人类对于物质世界的影响，却又往往充满了无知者无畏的情怀，以至于一点蝇头小利就可以让不同的族群世代为仇。石油危机引发的动乱可以说明这一点，这远比“贪婪”的后果更严重。

面对全球变暖的窘境，人类正在放下成见，无论是哪个国家还是哪个民族，都决定坐下来，重新认识物质科学的规律，一同探讨未来的出路。

在中国，雄心勃勃的“碳达峰”与“碳中和”计划正在执行当中。在这项计划中，到2030年，全中国排放的二氧化碳总量将达到最高值，自此以后，排放量会逐步降低；到2060年，全中国排放与吸收的二氧化碳将达到平衡。

为了实现这样的“双碳”目标，新的产业也在推进之中，它们不再依赖于石油这样的特定资源，也凝结了无数人的智慧。我们甚至可以说，它们是中国的工程，但是成果属于整个人类。



矿化固碳

在青海海西州，太阳能光电板正在铺设之中；在河北张家口，氢能源汽车开进了冬奥会场馆；在金沙江上，白鹤滩水电站续写三峡水

利枢纽的奇迹；在黄海边，全国最大的风电场张开叶片呼吸海风……

借助于这些科学技术，我们改变了人类对于物质惯有的利用方式。在获取有用资源的同时，减少甚至避免向环境中排放废物，让地球的物质循环依然保持着平衡。

“人猿相揖别，只几个石头磨过，小儿时节。”在天地之间，人是最特殊的生物，智慧的大脑让我们能够认识物质，同时审视自己。也正因为此，我们需要承担更多的责任，善待物质世界。

不再浪费，不再无知，这是我们与物质之间的全新关系，也是人类免于恐龙命运的唯一选择。

也只有这样，我们才能再一次发自内心地思考那个古老的问题：物质是什么？

# 9 物质是什么

回到我们畅游物质世界的起点，闭上眼睛想一想：物质是什么？

当我问起这句话的时候，就已经传递了一段信息，你感知到这段信息，并顺着我所提的问题，开始思考它。

我们之间的交流，本质是意识层面的——你需要理解我所说的文字。这一点，和信息所在的载体并无关系。换句话说，不管你现在是从纸张还是电子设备的屏幕上看到这段话，首先都需要在意识上去理解它，所以它和你的意识相关。相反，一个外星人，如果完全没有接触过地球上的文字，即便获取了同样的纸张或是屏幕，他也无从解读。事实上，人类早就遭遇过同样的困惑，我们至今还无法完整破译古埃及文字或甲骨文，以至于我们无法准确地知晓数千年前发生过的那些事。甚至只是一百多年前的历史，也留给我们很多空白，也许再也没有机会揭秘。



难以破译的甲骨文

另外一方面，我们总是会问起：恐龙是怎样灭绝的？这个话题经

久不衰，于是艺术家们借助于小说、电影之类的各种形式发挥想象，科学家们也不甘落后，找出6 500万年前小行星撞击地球的证据。我们并不怀疑，正是这场灾难，让恐龙走向灭亡，还好鸟类作为它们的后裔幸存至今。不止如此，还有数亿年前出现的生命形式，46亿年前的太阳系，还有138亿年前的宇宙，都是我们好奇的对象。

因为信息的缺失，我们很难和自己的祖先交流。但我们又如何跨过上亿年的层层阻隔，去和那些恐龙乃至非生命的天体或粒子交流？这实在是一件令人匪夷所思的事情。

诚然，我们总是希望更详细地弄清楚离自己更近的那些事，所以对于千百年来的这些记忆，我们总会觉得历史记载得太过简单，不足以解答所有疑问。至于那些比恐龙更遥远的故事，我们只想问个是非，根本不在意它们的编年史。

但这并非是本质缘由，我们的身体中就蕴藏着某些答案。

在传统的遗传学理论中，父母遗传给孩子的DNA，就决定了这个孩子未来的各种特质。然而，更进一步的研究却发现，事情远不是这么简单。

举个例子，当父母备孕之时，刚好遇上了饥荒。在这种条件下生育出来的孩子，成年之后会比那些和饥荒无缘的孩子更容易发胖。

研究人员猜测，父母将自己对饥荒的记忆以某种形式遗传给了孩子，尽管DNA中记录的遗传密码并没有什么差别，但是孩子却继承了这种对于饥饿的恐惧，对食物有着更高的热情，也更容易变胖。

也就是说，对于外界环境的反应，或许会以某种形式刻在我们的基因之中，并没有经历饥荒的孩子却在内心深处存在着对饥荒的记忆。如今，对这种可能性的思考，已经被纳入到方兴未艾的“表观遗传学”之中。

更进一步说，我们是否也拥有更久远的记忆，只是并不知道它们以何种形式左右着我们的意识？

这当然也是有可能的。

很多人都有过这样的体会：一个从未到达过的地方，或者一个从

未经历过的场景，当自己第一次身临其境之时，却感觉曾经来过，或是感觉曾经梦到过这一切。很遗憾，今天的科学界对此并无准确的解释。但我们几乎可以肯定，这不会是通灵，也不太可能会是时空穿越，而是有着属于这种现象的物质基础。

当我们把尺度拉到更长的时间轴上，还可以察觉更清晰的脉络：

作为无尾生物的人类，却有极少数个体会出现返祖现象，仿佛数百万年前的祖先那样长出尾巴——然而他们并没有携带尾巴生长所需的基因，至今我们还不知道这种远古记忆从何而来；

病毒被认为是地球上最原始的生命形态，也遍布整个地球，不管人类是否愿意，都不得不与它们结识，有时还会产生激烈的冲突——然而在人类的DNA中，存在着很多原本属于病毒的密码片段；

水是地球生命诞生的关键要素，单细胞生命经过数十亿年的演化成就了人类，而在人体中，占比最多的物质还是水，我们的祖先逐水而居——我们和水的亲密关系，是否早在单细胞生命的阶段就已经奠定？

构筑我们的各种分子，可以追溯到地球早期的地质运动。

构筑我们的各种元素，可以追溯到太阳系的形成之日。

构筑我们的所有物质，可以追溯到宇宙的诞生之时。

我们当然不会相信，宇宙中的那些简单粒子具备和人类一样的意识，但是在人体之中，它们的确又是我们人类意识的载体——这样看起来有些矛盾的表述，才真正揭示了物质世界的真谛。

从最基本的粒子，到只有一个质子和一个电子的氢原子，到一百多种不同的元素，到水和二氧化碳这样的小分子，再到蛋白质与核酸这样的复杂分子，物质经过漫长的发展，最终形成了生命物质，又从原始生命一路演化成人类。可以说，人类走到今天，就是一部物质的演化史，我们的身体早就打上了物质演化的烙印。这些烙印，并不会因为时间久远而消散。因此，我们不只是会对祖先创造的文明感兴趣，还会花费精力去探索物质的起源——那也是我们自己的起源。

物质，就是我们自己。

# 主要参考文献


[1]拉瓦锡．化学基础论[M]．任定成，译．北京：北京大学出版社，2008．

[2]达尔文．物种起源[M]．舒德干，译．北京：北京大学出版社，2005．

[3]阿尔伯特·爱因斯坦．狭义与广义相对论浅说[M]．张卜天，译．北京：商务印书馆，2018．

[4]史蒂芬·霍金．时间简史[M]．许明贤，吴忠超，译．长沙：湖南科学技术出版社，2010．

[5]莱纳斯·鲍林．化学键的本质[M]．卢嘉锡，黄耀曾，曾广植，等译．北京：北京大学出版社，2020．

[6]埃尔温·薛定谔．生命是什么[M]．张卜天，译．北京：商务印书馆，2018．

[7]曹天元．上帝掷骰子吗[M]．沈阳：辽宁教育出版社，2006．

[8]马克·米奥多尼克．迷人的材料[M]．赖盈满，译．北京：北京联合出版社，2015．

[9]施普林格·自然旗下的自然科研．自然的音符：118种化学元素的故事[M]．Nature自然科研，编译．北京：清华大学出版社，2020．

[10]吉姆·巴戈特．完美的对称[M]．李涛，曹志良，译．上海：上海世纪出版集团，2012．

[11]瓦茨拉夫·斯米尔．能量与文明[M]．吴玲玲，李竹，译．北京：九州出版社，2021．

[12]尼尔·德格拉斯·泰森，唐纳德·戈德史密斯．140亿年宇宙演化全史[M]．阳曦，译．北京：北京联合出版社，2019．

[13]孙亚飞．元素与人类文明[M]．北京：商务印书馆，2021．







# 给青少年讲生命科学

刘锐 著

清华大学出版社

基础前沿科学史丛书

# 给青少年讲生命科学

刘锐　著

清華大学出版社
北　京

图书在版编目（**CIP**）数据

给青少年讲生命科学 / 刘锐著．—北京：清华大学出版社，2022.10

（基础前沿科学史丛书）

ISBN 978-7-302-61943-7

Ⅰ．①给…　Ⅱ．①刘…　Ⅲ．①生命科学－青少年读物　Ⅳ．①Q1-0

中国版本图书馆CIP数据核字（2022）第180870号

责任编辑：胡洪涛

封面设计：意匠文化·丁奔亮

责任校对：王淑云

责任印制：宋林

出版发行：清华大学出版社

网　　址：http://www.tup.com.cn, http://www.wqbook.com

地　　址：北京清华大学学研大厦A座

邮　　编：100084

社总机：010-83470000

邮　　购：010-62786544

投稿与读者服务：010-62776969，c-service@tup.tsinghua.edu.cn

质量反馈：010-62772015，zhiliang@tup.tsinghua.edu.cn

印装者：三河市龙大印装有限公司
经　　销：全国新华书店
开　　本：165mm × 235mm
印　　张：10.5
字　　数：115千字
版　　次：2022年11月第1版
印　　次：2022年11月第1次印刷
定　　价：55.00元

---

产品编号：097596-01

# 丛书序　给面向青少年的科普出版点一把新火

2022年是《中华人民共和国科普法》通过的第20年，在这样一个对科普工作意义不凡的年份，由北京市科学技术委员会（以下简称市科委）发起，清华大学出版社组织的“基础前沿科学史丛书”正式出版了。这套书给面向青少年的科普出版点了一把新火。

2022年9月4日，中共中央办公厅、国务院办公厅印发《关于新时代进一步加强科学技术普及工作的意见》，进一步强调“科学技术普及是国家和社会普及科学技术知识、弘扬科学精神、传播科学思想、倡导科学方法的活动，是实现创新发展的重要基础性工作”。科学技术普及是科技知识、科学精神、科学思想、科学方法的薪火相传——是“薪火”，也是“新火”。

市科委搭台，出版社唱戏，这套书给面向青少年的科普图书出版模式点了一把新火。市科委于2021年11月发布了“创作出版‘基础前沿科学史’系列精品科普图书”的招标公告，明确要求中标方在一年的时间内，以物质科学、生命科学、宇宙科学、脑科学、量子科学为主题，组织“基础前沿科学史”系列精品科普图书（共5册）出版工作；同步设计制作科普电子书；通过网络媒体对图书进行宣传推广等服务内容。这些服务内容以融合出版为基础，以社会效益为初心。服务内容的短短几句话，每一句背后都是特别繁复的工作内容。想在一年的时间内，尤其是在2022年新冠肺炎疫情期间，完成这些工作的难度可想而知，然而秉承“自强不息，厚德载物”的清华大学出版社的出版团队做到了。

中国科学家，讲好中国故事，这套书给面向青少年的科普图书选题内容点了一把新火。中国特色社会主义进入新时代，新一轮科技革命和产业变革正在深入发展，基础前沿科学改变着人们的生产生活方式及思维模式。《中华人民共和国国民经济和社会发展第十四个五年规划和2035年远景目标纲要》提出：在事关国家安全和发展全局的基

础核心领域，制定实施战略性科学计划和科学工程。物质科学、生命科学、宇宙科学、脑科学、量子科学等领域，迫切需要更多人才参与研究，而前沿科学人才的建设培养，要从青少年抓起。这5本书的作者都是中国本土从事相关专业领域工作的科学家，这5本书都是他们依托自己工作进行的原创性工作。虽然内容必然涉及科学史的内容，但中国科学家尤其是近些年的贡献也得到了充分展示。

初心教育，润物无声，这套书给面向青少年的科普图书科普创作点了一把新火。习近平总书记提出：科技创新、科学普及是实现创新发展的两翼，要把科学普及放在与科技创新同等重要的位置。因此，针对前沿科技领域知识的科普成为重点。如何创作广受青少年欢迎的优秀科普图书，充分发挥科普图书的媒介作用，帮助青少年树立投身前沿科学领域的梦想，是当前科普出版工作的重点之一，这对具体的科普创作方法提出了要求。这套书，看得出来在创作之初即统一了整体创作思路，在作者进行具体创作时又保持了自己的语言习惯和科普风格。这套书充分体现了，面向青少年的科普图书创作，应该循序渐进，张弛有度，绘声绘色，娓娓道来，以科学家的故事吸引他们，温故科学家的研究之路，知新科学家的科研理念，以科学精神润物细无声。

靡不有初，鲜克有终。2022年10月16日，习近平总书记在中国共产党第二十次全国代表大会报告中强调“教育、科技、人才是全面建设社会主义现代化国家的基础性、战略性支撑”。且将新火试新茶，诗酒趁年华。希望清华大学出版社的这套“基础前沿科学史丛书”为广大青少年推开科学技术事业的一扇门，帮助他们系好投身科学技术事业的第一粒扣子，在全面建设社会主义现代化强国的新征程上行稳致远。

中国工程院院士

清华大学教授



# 前言

科学探索是一个不断追求真理的过程，且追求科学真理需要具有质疑和批判的精神，当然这必须建立在严谨认真、实事求是的基础之上。同时，知识无尽，学海无涯，科学研究是最为典型的、高级的、创造性的人类活动，需要一代又一代的有志青年前赴后继地投身其中。

质疑和批判是为了去伪存真，是为了让科学沿着一个正确的方向不断前行。正如“进化论之父”达尔文所说：“无知有时比知识更容易带来自信，正是那些所知甚少的人，而不是那些学识渊博的人，会如此肯定地断言这个或那个问题永远无法被科学解决。”因此，我希望能尽一点微薄之力，将自己知道的一些生物学知识介绍给广大青少年读者，让大家了解生命科学史上的重大发现及其背后不一样的故事。

根据史料记载，人类对生物的研究已有2000多年的历史。然而，直到19世纪初，“biology”（生物学）一词的诞生，生物学才拥有了自己的“铭牌”。无论是在古代的欧洲国家，还是在古代的中国；无论是医学、解剖学，还是农学、植物学、动物学，都属于生物学的研究范畴。当前，生物学的发展可以说是一日千里，重大研究成果不断涌现，深刻地影响着人类的社会发展和文明进步。尽管如此，依然存在着大量的未知之谜，有待我们去深入探索。

科学著作及其阐释的内容在本质上都是为了呈现科学真相和传承科学文明。本书面向广大青少年读者，深入浅出地介绍了包括生命起源的假说、DNA双螺旋构建的过程、埃博拉病毒的暴发、基因编辑技术的发展等生物学史上的重大事件或前沿成果。我希望本书能够为广大青少年读者带来一些启发，以便大家能够体察悟理、见微知著，从小在心中埋下一颗科学、理性的种子，让热爱科学、探索未知的精神在心中萌芽、开花。

也许，我们并不一定都能成为明日的科学巨匠，但是至少可以让求真务实的科学思想和开拓创新的科学精神照亮我们未来的人生旅

途！

# 目 录

# 1　在毒与火的深渊——生命的起源

在进行后续的讲述前，首先要了解生命科学中最为重要的一个话题，那就是我们生命的起源。这是人类自出现以来一直在思考却又没有得到完美解释的问题，也是探讨生命科学的基础。

面对地球长达40多亿年的漫长演化历史，生命的诞生及生命诞生之初的环境对我们来说都是遥远、神秘而又无法直观感受到的。要还原生命起源的真相，不仅需要多学科的合作，更需要大胆地假设和小心地求证。



地球的诞生

可以说，我们始终没有停下探索的脚步，那么生命究竟缘起于何？迄今为止，众说纷纭，出现了很多种不同的理论，但是始终没有一个完全令人信服的答案，现在就让我们共同来了解一下究竟有哪些关于生命起源的假说。

## 1.1　笼罩世界的神创论

放射性元素的半衰期表明地球的年龄约为46亿年。在这么漫长的历史中，生物究竟是如何诞生的？又是从何时开始诞生的？这是让人着迷却又难以得到完美解释的话题。在科技不发达的古代，人们都渴

望了解，大自然中千奇百怪的生物都来自哪里？是神创造的？是自然产生的？还是由其他物种演变而来的？这些问题曾深深地困扰着人们。



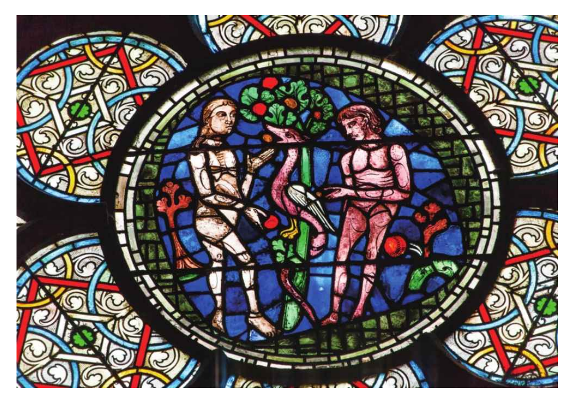
女娲造人

在近代科学诞生之前，神创论占据着重要的地位。当时人类没有办法用科学理论来解释大自然中的种种神秘现象，于是就只能借助于神的力量。当时人们普遍认为，世界上的所有生命都是由神创造的！

《圣经》中曾经描述了这样的场面：上帝在6天的时间里，先后创造了日月星辰、山脉河流、树木花草、飞禽走兽。然后上帝依照自己的模样创造了亚当，又用亚当身上的一根肋骨创造了夏娃，夏娃和亚当共同在伊甸园中生活。

我们也经常把神创论称为特创论。神创论的一个重要特点就是，所有已经创造好的物种是不会再发生变化的，即使有所变化，那也是在很小的范围内发生一些改变，而不会变成另外的物种。

在科技并不发达的年代，对于眼前的世界，人类缺少最基本的认知，大家对于自然现象的伟大和神奇，充满了羡慕和崇敬，为了能够合理解释这些自然现象，人们只能大胆地幻想。而这种上帝已经安排好了所有的剧本、地球上的所有生命体只要按照剧本的要求去演绎就行了的假说，成为当时人们的不二选择。

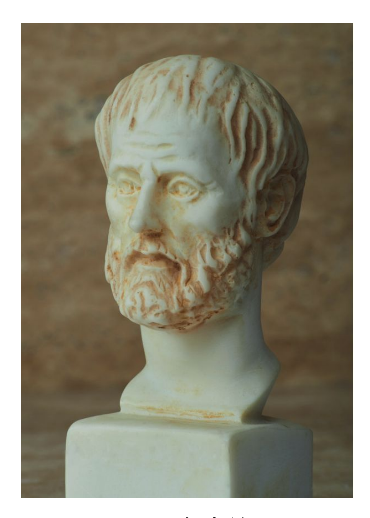


法国巴黎圣母院染色玻璃上的亚当、夏娃

## 1.2　自然发生的小白鼠

在科技不发达的古代，关于生命起源，除了神创论之外，还萌生了很多在现在看来十分可笑的说法，然而，这些错误观点却深深地影响了学术界很多年，甚至这些观点的支持者中不乏亚里士多德（Aristotle）和牛顿（Newton）这样的科学巨匠。

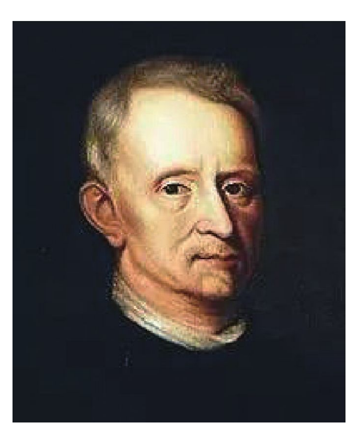
当时有一种自然发生论，它最核心的观点是：生命，尤其简单的生命，是由无生命的物质自然发生的。最早支持自然发生论的人是亚里士多德。作为当时科学界的权威人物，他的态度决定了很多人对这些问题的看法，毕竟绝大多数人都会选择相信科学巨匠的判断。事实却多次证明，学术大师们在某些问题上的看法未必就是正确的，大家应该独立地思考，而不是一味地盲从权威。



亚里士多德

虽然学术大师的光环在某些时候可以促进学术传播，但是在另一些时候也会阻碍发现真理的进程。在生命起源问题上，亚里士多德就成为了阻碍科学发展的人。

亚里士多德认为物质是自然发生的，甚至还给各种物质的来源编制了一个目录。他认为，每一种物质的繁殖都需要“热量”，这种热量是最关键的。高等动物是通过“动物热”产生的，低等动物是在雨水、空气和太阳热的共同作用下从黏液和泥土中产生的。例如，晨露同黏液或者粪土在一起反应就会产生萤火虫、蠕虫、黄蜂……而黏液会自然产生蟹类、鱼类、蛙类，老鼠则是从潮湿的土壤中产生的。在现在看来，这些观点违反了基本的科学常识，十分可笑。但是在当时，这些观点却被认为是普适的真理。牛顿也曾为自然发生论摇旗呐喊，他认为植物是由逐渐变弱的彗星的尾巴形成的，这种说法让人大跌眼镜，甚至觉得难以想象，发现了力学三大定律的牛顿，怎么会有这样的认知？因为他们在学术上有着巨大成就和影响，所以有了他们的支持，自然发生论就有了更加广阔的市场。



海尔蒙特

当时，还有很多著名的科学家都支持自然发生论。因为在肮脏的环境中容易发现老鼠和苍蝇，所以很多人想当然地认为，老鼠和苍蝇是在肮脏的环境中自然产生的。著名的科学家海尔蒙特（Helmont）就曾经提出：把糠和破布塞进一个瓶子里，将瓶子放在阴暗的床底下就会生出来小老鼠。海尔蒙特是17世纪著名的化学家和哲学家，他是引导炼金术向化学学科转变过程中的重要人物，也是最早发现二氧化碳（$CO_2$）的人。他认为木头等物质燃烧后得到的是野气，也就是我们常说的二氧化碳。海尔蒙特在化学方面的工作是突破性的，但是在生物自然发生论上却摔了一个大跟头。仔细分析他的这些说法，可以发现，其中有很多漏洞。例如，如何确认这些老鼠或苍蝇不是从外界进入的呢？这些实验的环境不是完全封闭的，即使是封闭的，也不能排除这些没有经过消毒的肮脏的破布中原先就存在着苍蝇卵，在合适的温度之下，这些卵很有可能会孵化出蛆来。

当我们现在掌握了科学知识，再回过头来看这样的说法，就知道它们是站不住脚的，但是，我们也不能一棍子将这些观点全部打死，我们应该把这些说法放在当时的历史条件下去看，在科技极其落后的古代能够进行这样深入的思考，提出这样的观点，也是有一定贡献的。

## 1.3 “原始汤”中的生命

对于生命的起源，还有一种化学起源假说，相比自然发生论，这一假说更容易获得多数人的认可。该假说的主要内容是，生命的诞生经历了两个主要阶段：第一阶段，在原始的大气和海洋中，发生了前期的化学反应，形成了最原始的有机物和生物大分子体系；第二阶段，这些大分子形成原始的生命。

化学起源假说中的原始地球海洋，充满了各种不定的因素，里面具备了化学反应发生的各种条件，也有大量的小分子物质，就像一锅诞生生命的“原始汤”。“原始汤”的理论是由苏联生物化学家亚历山大·伊万诺维奇·奥巴林（Oparin, Alexander Ivanovich）和印度生物学家霍尔丹（Haldane）共同提出的。而我们教科书中经常提及的“原始汤”实验是由奥巴林所创造的。



原始海洋的图景

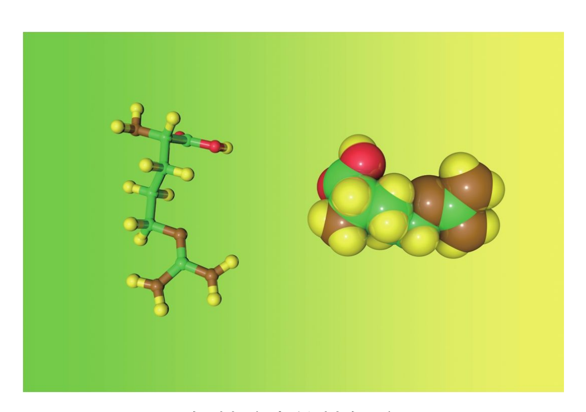


原始汤

奥巴林在1936年出版了《地球上生命的起源》，在该著作中，他提出了关于生命起源的假说。他认为在原始的地球大气中，充斥着大量的宇宙射线、紫外线、闪电等一些蕴含着大量能量的能量源，在它们的作用下，原始大气中的$CO_2$、$N_2$、$H_2S$、$H_2$、$NH_3$等不断地发生着聚合反应，最终形成小分子化合物，如氨基酸、嘌呤、嘧啶、核糖等，这些都是构成大分子生命物质的基本成分。

原始的地球就像是一个巨大的反应容器，内部火山持续喷发，温度很高，很多由火山喷发带来的气体，包括$CH_4$、$NH_3$、HCN、$H_2S$、$CO_2$等共同组成了原始的大气，在强烈的紫外线、宇宙射线、高能粒子流、闪电等能量的作用下，合成了各种小分子化合物。

为了进一步验证这个假说的正确性，科学家们尝试在实验室里模拟原始地球中的大气和海洋环境，看看能不能在这些能量源的作用下将无机物合成简单的生命基本物质，包括核酸和蛋白质的前驱体：氨基酸、多肽、核苷酸……



氨基酸中的精氨酸

1953年，美国芝加哥大学的研究生米勒（Miller）在导师的指导下，将水注入500毫升的烧瓶，同时把玻璃瓶中的空气抽走，加入模拟原始地球还原性大气的由$CH_4$、$NH_3$、$H_2$等组成的混合气体。实验中，他持续加热烧瓶，让水蒸气和模拟原始大气中的混合气体通过密闭的管道进入另外一个容量为5升的大烧瓶中，接着通过电火花放电来模拟原始大气中的雷电。一周之后，实验团队检测聚集在容器底部

的溶液，看有没有简单的物质生成，这些溶液就相当于生成的物质被雨水冲淋后形成的原始海洋。经过化学检测，结果发现，这些经过放电冷却的溶液中存在20种有机物，其中包括11种氨基酸，而这11种氨基酸中有4种氨基酸是人体必需的氨基酸，它们分别是甘氨酸、丙氨酸、天冬氨酸和谷氨酸。



蓝色闪电弧放电

这个实验的结果证实了奥巴林和霍尔丹假说的正确性。米勒对假说中还原性大气的成分进行了一些改动，他认为原始地球中还原性大气的主要成分应该是$CH_4$、$N_2$，以及微量的$NH_3$和$H_2O$，因为大量的$NH_3$会直接溶于水中。随后米勒和同事又做了大量的实验，通过不断地变换实验条件，分别利用紫外线、β射线、高温等作为能源，同时更换混合气体的成分，例如，用$H_2S$替代$H_2O$，用HCN替代$CH_4$，实验的结果表明都能够产生氨基酸小分子。

氨基酸小分子产生之后，生命的起源就有了最基本的原料。如果在合适的环境下，例如在海底火山口附近的高温条件下，就可以进一步发生缩合反应，生成一些氨基酸的聚合体。后来有学者也进行了相应的实验，把20多种天然氨基酸按照酸性、碱性、中性分别进行混合，加热到170摄氏度左右，就可以得到某些类蛋白物质。

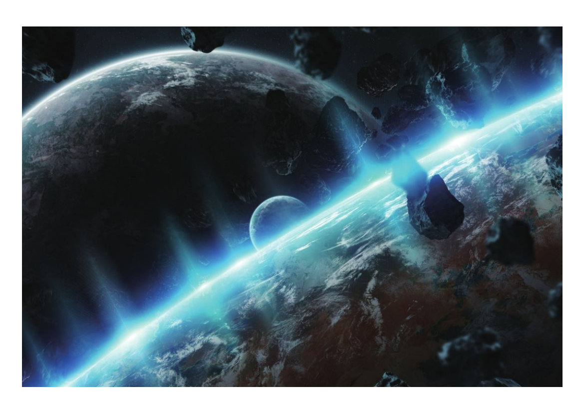
这个假说其实也得到了证实，在原始的大气环境中，通过适当的能量源的作用，是可以生成有机物和生物大分子的。从无机到有机、

从简单到复杂的化学反应在原始地球的环境下是可以完成的，但是能不能由这些大分子演化成简单的生命依然没有定论。

## 1.4 来自外太空的礼物

除了上面我们说到的几种生命起源的观点之外，还有一种比较流行的观点：生命起源于外太空，也就是常说的“泛生假说”，又被称为“宇宙胚胎种源假说”。

地球的年龄大约为46亿年，这样漫长的时间对于人类的历史来说太长了，但是地球对于浩瀚的宇宙来说，又显得那么微不足道，因此在广袤的太空中可能存在着更多的未知和生命起源的可能。



外太空中的陨石

1907年，瑞典化学家斯万特·奥古斯特·阿列纽斯（Svante August Arrhenius）最先提出了“宇宙胚胎种源假说”。他在著名的《宇宙的形成》一书中提出了这样一种观点：在广袤的太空中漂浮着大量的“生命胚种”，这种所谓的“生命胚种”就是生命最原始的形式，可以发展诞生出生命。它们在太空中随着太阳风、黑洞压力等外力的作用四处飘荡。有机会伴随着陨石、彗星、星际尘埃等降落到一些星球的表面，如果恰好这个星球有合适的条件进行生命的孕育，那么就会诞生出最原始的生命。这样外太空的“生命胚种”就从一个星球传播到了另外一个星球。而地球上的生命正好来源于外太空的“生命胚种”，是外太空的礼物，“生命胚种”造就了地球上繁盛而多样的物种。

这种理论乍一听有点儿脑洞大开，但是它自从诞生伊始，就得到很多科学家的支持，包括美国国家航空航天局（National Aeronautics and Space Administration, NASA）的天体物理学家，他们认为地球上的生命很可能是起源于40亿年前坠入海洋的一颗或者数颗彗星。



太空中的陨石

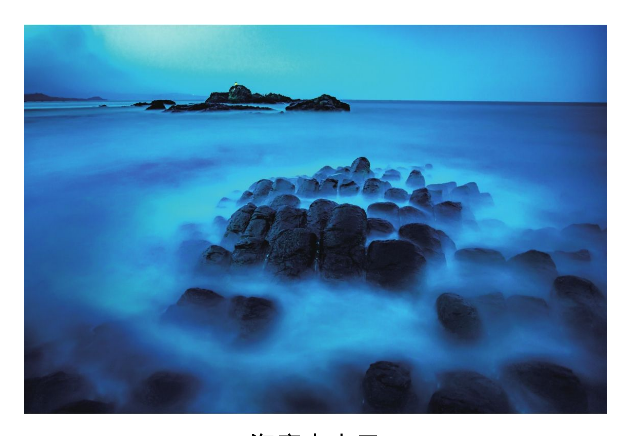
目前，地球上也存在很多从外太空坠落下来的陨石，其中被研究得最为广泛的一颗是1969年坠落在澳大利亚默奇森镇的被命名为“默奇森”的陨石。这颗陨石中含有70多种氨基酸，包括常见的甘氨酸、谷氨酸、丙氨酸等，还包括两种构成生命不可缺少的核酸分子：尿嘧啶和黄嘌呤。而且经过放射性碳测年分析，证实了这些分子都是在外太空就已经形成了，给“宇宙胚胎种源假说”提供了最直接的证据。

但是我们仔细思考一下，似乎这样的假说存在着一些难以解释的问题。例如，在浩瀚的外太空中，存在着极具破坏性的射线，如紫外线等，它们对于生命物质，包括所说的“生命胚种”都有强烈的致死性，那么这些原始的生命物质是如何存活下来的呢？另外，在进入地球的过程中，还要经历一系列高温、高热的严酷环境的考验，地球上当时的生存环境也未可知，这些简单的蛋白质分子或者核酸分子能够在这样的环境下保持活性吗？因此，这种说法还有待进一步的证实。

## 1.5　在毒与热的深渊

目前，关于生命的起源，还有一种假说得到了更多人的认可。这种假说认为生命应该是起源于原始的海洋底部。

1979年，在太平洋中深达2000米的海底，科研人员发现了很多冒着黑色溶液的喷发口。这些黑色的溶液在被冰冷的海水冷却之后，迅速地堆积形成一种独特的柱状体。这些柱状体附近存在着很多原始的微生物群落，包括嗜硫细菌、古细菌等，它们可以通过氧化硫离子、锰离子、亚铁离子等物质来获取能量，将无机碳转化为有机碳。而海底涌出的丰富矿物元素也给这些细菌的繁殖和生长提供了充足的养分，这样就形成了一个特殊的生态系统。



海底火山口

海底喷出的液体，温度高达350摄氏度，并且含有$H_2S$、$CH_4$、CN等小分子，为非生物有机合成提供了条件。这些化能自养的细菌可以利用热泉中喷出的硫化物中的能量去还原$CO_2$，来制造有机物。

地质学的研究表明，海底热液喷口附近的环境条件和原始地球的环境极其相似，因此，很多科学家就猜测原始的生命很可能就是起源于海底的水热环境中，而不是原先认为的原始大气。那么在这样的环境中，就有了孕育生命的可能，因为在这样的环境中，存在着大量丰富的还原性物质，可以通过氧化这些还原性的物质来提供能量，然后逐步地将无机物变成有机物，再从小分子合成大分子，继而逐步地孕育出生命。

这种假说得到了来自基因组测序结果的支持和地质学研究的支持。美国伊利诺伊大学的研究人员卡尔·沃伊斯（Carl Woese）等对海底热液口的细菌进行了基因组序列分析，得出结论：这些细菌是最简单、最古老的古细菌原核生物，与真细菌和真核生物并列为第三界，足以说明它的古老历史。1996年，美国基因组学研究所的研究人员也对太平洋海底的一种产甲烷的细菌进行了基因组序列分析，破译了1700个基因密码，确定了这类古细菌是与真细菌和真核生物不同的第三种分支，而且从起源的角度来说更为古老，借此推断这些古细菌是原始生命最早的形式。

另外，地质学上的证据也间接地支持了这一假说。我们知道，在没有经受变质作用的古老岩石上会保存着生命活动所遗留下来的痕迹，这也是地质学中探索生命起源的重要依据。1997年，在格陵兰的伊塔地区发现了被认为是极其古老的岩石，该岩石经过分析已经有38.5亿年之久，也是迄今为止所发现的最为古老的海洋沉积物，在岩石中有明显的生物存在的痕迹，而在这么古老的年代就有生命活动的痕迹，也间接说明了生命起源于原始海洋底部的古细菌。



甲烷八叠球菌属古生菌

这种假说提出之后，也遭到了很多人的质疑。例如，有人认为在海底附近的高温会对化学反应的进行起到促进作用，但是高温也对组成生命体的蛋白质有着极强的破坏性；形成的黑色的柱状体复合物寿命很短，只有几十年的时间，这么短的时间内是无法孕育出生命的，

并且周围的环境酸性太强，pH值为1～2，不利于生命的诞生；等等。

在生命起源多种假说的“百家争鸣”之下，还出现过几次影响较大的科学争论。第一次是在自然发生论与生源论之间展开的。生源论是相对于自然发生论而言的，由法国生物学家路易斯·巴斯德（Louis Pasteur）所提出，即生物不能够自然发生，只能通过生物的繁殖产生。最终，巴斯德通过一系列精确实验提供了确凿的证据，宣告生源论的彻底胜出。第二次争论是宇宙胚胎种源假说与化学起源假说的争论，由于在外太空广泛存在的紫外线和宇宙射线对于生命物质具有强烈的杀伤作用，因此化学起源假说占据了上风。第三次争论的主角也是这两种理论，以奥巴林学说为代表的原始地表化学起源说已基本成熟并迅速推广开来，可是半路杀出个程咬金，伴随着射电天文学和宇宙化学的飞速发展，宇宙中存在的大量有机物相继被发现，宇宙胚胎种源假说又占据了上风。



生命的起源

面对众说纷纭的假说，生命起源的真相究竟是什么，依然是一个未知数，也需要等待我们进一步的研究。但是我们相信，终究有一天，我们能够顺利地揭开它的神秘面纱！



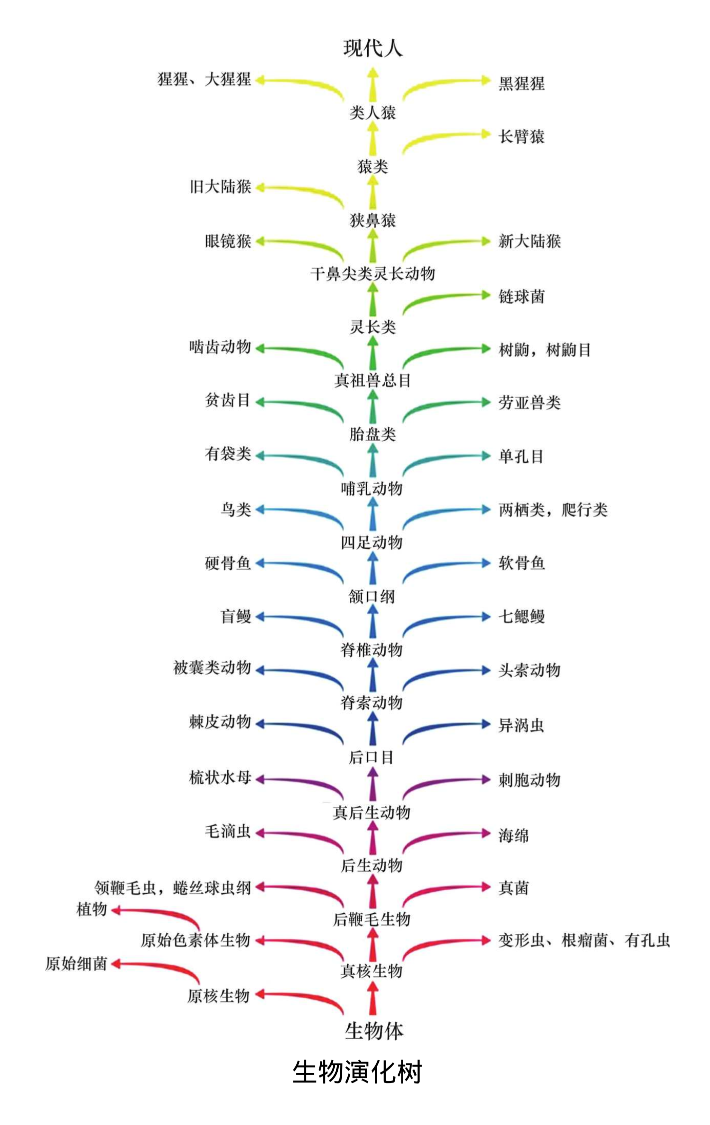
# 2 是“进化”还是“演化”

在生命诞生之后，地球上的物种便开始了长达几十亿年的演变之路，最终形成了丰富多样的物种群体。

由于不断有新的物种诞生，也有老的物种灭绝，在人类文明发达之前，我们无法对所有物种一一记录，因此没有办法确认自从生命诞生以来全世界究竟演化出了多少物种。可以肯定的是，绝大多数的物种已经灭绝了，现存的物种数量只是生命诞生以来所有物种数量的一小部分。

## 2.1 动植物该如何分类

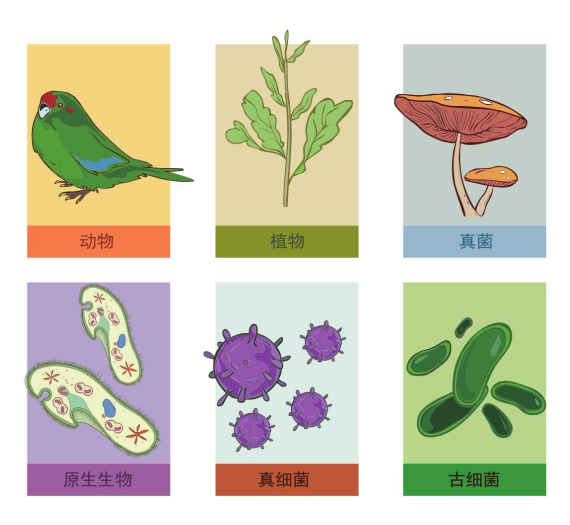
如果我们要认识和了解这些物种，就必须对它们进行系统的分类。早在文字出现之前，人类就开始对生物进行分类。很多博物学家开始将具有同种性征的物种归类在一起，按照从无到有、从简单到复杂的顺序将它们一一排列起来。这样就会让这些博物学家萌发一个想法：生物究竟是在什么时间、什么地点起源的，其中有没有某种内在的联系呢？哪一种生物是现在所有生物的祖先呢？




生物演化树

在古代欧洲，一直按照柏拉图（Plato）提出的两叉式分支法来划分动物的种类。例如，把动物分为水栖动物和陆地动物，有翅膀的动物和无翅膀的动物等，这是一种有着明显对立特征的动物分类方式。这种简单的分类方式虽然有一定的道理，但是却存在着致命的缺陷，容易人为地造成同一物种的分裂，让人明显地感觉到这种分类方式是不正确的。例如，我们把有翅膀的蚂蚁分类在有翅膀的动物中，把没有翅膀的蚂蚁分类在无翅膀的动物中，人为地把蚂蚁这一物种分割开

来，显然是不合适的。因此，在这种分类方式出来之后，不少学者表示了质疑。




简单的分类方式

柏拉图的学生亚里士多德萌生了初步的分类学思想，认为可以找到更加合理的分类方式。他描述了500多种动物，并对这些动物进行了分类。他按照动物有无红色的血液将动物分为有血动物和无血动物两类。虽然在现在看来，这种分类方式过于简单，但是在当时，这种分类方式还是具有重要价值的。其实我们现在知道，依据动物的血液进行分类是一件相对来说比较复杂的事情。一些低等动物，如原生生物界的原生动物水螅、涡虫、绦虫、蛔虫等，因为它们的肌体没有高度分化，通过体液渗透就能够满足相应的循环供氧需要，所以它们是没有血液的。另外，生物血液的颜色也并非都是红色的。肢口纲的鲎在氧合状态下血液为蓝色，在非氧合状态下血液为无色或白色；有些多毛虫，如帚毛虫科、绿血虫科动物的血液在氧合状态下为红色，在非氧合状态下为绿色；腕足类动物的血液在氧合状态下为紫红色，而在非氧合状态下为褐色；虾、蜘蛛、乌贼等动物的血液是青色的；节肢类动物的血液是无色或淡蓝色的……亚里士多德认为红色的血液才是血液，而具有其他颜色血液的动物在他的理论里就会被归类为无血动物。

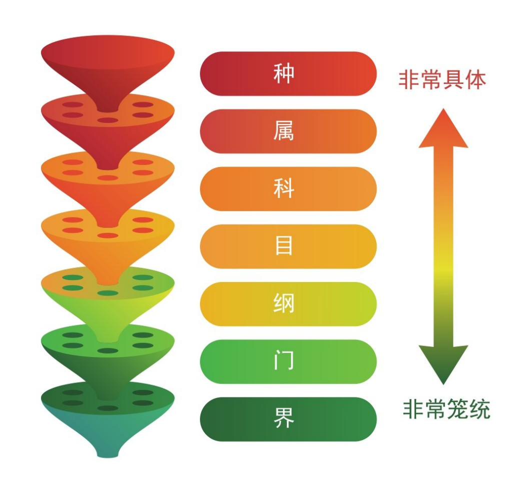


鲎

亚里士多德是欧洲第一位创立动物学分类理论的学者，也是第一位按照动物性状特征进行动物分类的学者。同时，他在植物分类方面也做了大量工作，由于种种原因，他的研究成果没有被保存下来，但是他的学生、植物学家狄奥弗拉斯图（Theophrastos）明确地区分了动物、植物，阐明了两者之间的区别。他提出了一个很有意思的观点：植物体在损失一部分身体后，会很容易得到更新，而动物体在失去一部分身体后，更新是极其有限的。这成为区分动物、植物的重要特征之一。

整个自然界的生物物种数量是极其丰富的，目前已知的生物物种数量约为200万种，而已经灭绝的生物物种数量则高达1500万种。因此，如果没有一种公认的分类方式，各个研究机构或学者自说自话，那么，不仅会导致大量的重复研究和资源浪费，还不利于信息的传递与交流。瑞典博物学家卡尔·冯·林奈（Carl von Linne）是生物分类学的先驱，他用了5个月时间进行野外考察，采集了大量的植物标本。通过实地考察，他对植物标本进行了分类整理，按照相似的形态特征进行编组，并在心中开始思考酝酿，什么样的特征才是整个植物界的分类标准。1735年，林奈在荷兰获得了博士学位，并出版了他的第一本博物学著作——《自然系统》。在自然界中的动物如猴子、鹰、苍蝇等，虽然在生活中人们会以统一的名称来称呼它们，但是在研究中，每个名称下可能还包含着多个物种，同一个物种在各地还有着不同的称谓。所以制定统一的命名法则成为当务之急。林奈提出，对自然界的物种应该按照界、纲、目、属、种的分类方式进行系统的分

类，同时提出了双名制命名法的命名法则。第一个名字是属名，后一个名字是种名。例如，人类的学名Homo sapiens就是林奈制定的，其中Homo是人属，sapiens是智慧的意思，所以可以称为智人。林奈首次将生物学中的物种分到一个多级的分类系统中。每一级成为一个分类阶元并沿用至今，现在已经逐步完善成为7个基本的阶元，其从大到小的顺序是：界、门、纲、目、科、属、种。其中还可以添加一些子单元，如在目下增加一个亚目，在科上增加一个超科。分类学按照界、门、纲、目、科、属、种的方式对自然界中的动物、植物进行系统的分类，按照器官的相似性，将类似的生物归纳在一起，形成一条完整的生物进化谱系。



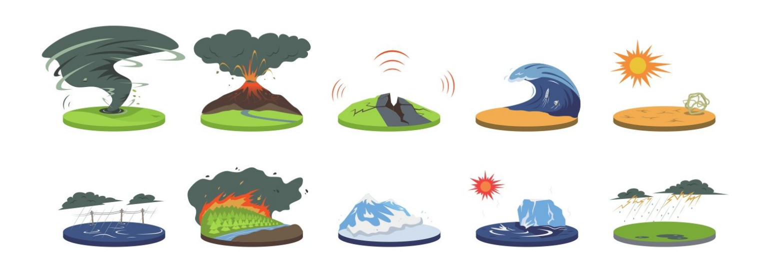

分类系统的等级

## 2.2 百家争鸣

在生物体按照特定的性状被归类为同一物种之后，我们可以清楚地看到同一物种的生物从简单到复杂的演化关系，这不禁让人产生思考，按照这样的思路追本溯源，同种生物的祖先是不是起源于同一属，同一属的生物是不是来源于同一科，同一科的生物是不是来源于同一目？最终，所有的生物是不是有一个共同的祖先呢？这个祖先又会是什么生物呢？

在进化论发表的前夜，社会上充斥着各种思想。例如，德国魏尔纳（Werner）的水成论、英国杰姆斯·赫顿（James Hutton）的火成论、法国乔治·居维叶（Georges Cuvier）的灾变论、英国查尔斯·赖尔（Charles Lyell）的地质渐变论等，其中，居维叶的灾变论和赖尔的地质渐变论最具影响力。居维叶是法国著名的博物学家，也是介于拉马克（Lamarck）和达尔文（Darwin）之间的一位划时代的人物。灾变论是地质学史上的一项重要理论。灾变论并不是居维叶首先提出来的，在他之前已经出现了很多不同种类的灾变论。

居维叶的理论其实并不新鲜，17—18世纪涌现出的大量灾变假说为他的理论奠定了基础。当时法国有一位著名的学者博内（Bonnet），提出了一个观点：世界会发生周期性的大灾难，每次灾难都会毁掉地球上存在的一切生物，然后又会重新创造出比之前更为高级的生物。他甚至还预言，在未来的某一次灾变后，在猴子和大象中会出现一个培罗（Pelor），在海狸中会出现一个牛顿或者莱布尼茨（Leibniz）。这是典型的灾变学说。居维叶提出了自己的理论——灾变论：世界经历了多次大的灾难，例如洪水，大规模的洪水将世界上的一切生物都毁灭了。在毁灭了所有的生物之后，造物主又创造出新的生命。他的观点像是进化学说与宗教学说的结合体。自然界确实发生过很多次大范围的灾难，包括导致恐龙灭绝的大灾难。灾变事件的存在是可信的，然而居维叶认为灾难之后，是造物主创造了新的生命，这就又回到了唯心主义的观点上。



不同的自然现象

赖尔的地质渐变论也有着重要的影响。赖尔是一位坚定的进化论拥护者，他在对火山的研究中发现，地质的变化是渐变的，是长时间

累积的过程，是经过上亿年自然力的作用后逐步形成的。他的著作《地质学原理》多次再版，他以优美的笔调将进化思想广泛传播，为进化论的诞生奠定了坚实的基础。



赖尔

赖尔出生于苏格兰福法尔郡金诺地村，17岁时进入大学学习，并痴迷于考察地质和采集化石，他在学校里参加了地质考察组，到处参观考察。经过大量的考察实践，他对地质学产生了浓厚的兴趣。凑巧的是，赖尔与灾变论以及火成论都有很深的渊源。赖尔的老师巴克兰（Buckland）是一位忠实的灾变论的粉丝，他对居维叶有着极强的个人崇拜，因此在讲课中掺杂了大量的个人情感。然而，赖尔却不为所动，他在自己的著作中表达了对于灾变论反对者的同情，因此不可避免地与老师产生了学术分歧。赖尔最喜欢的一本著作是科学家普雷菲尔（Playfair）的《关于赫顿地球论的说明》，赫顿是火成论的创立者。火成论的主要观点是：花岗岩的矿物晶体结构不可能是水中沉淀的产物，而是岩浆冷却后的结晶物；花岗岩脉与其他层状岩石的穿插切割关系，也说明它不是沉积的而是地下岩浆活动的结果。赖尔对这种朴素的唯物主义观点有着发自内心的强烈认同感。

19世纪20年代，赖尔开始了他的地质考察之旅，他的足迹遍布了英国、法国、瑞士、意大利、德国等国家。他在这次考察中有一项重要使命，就是为自己的著作《地质学原理》寻找实际物证。在考察过程中，他有幸结识了拉马克（Lamarck）、居维叶、亚历山大·冯·洪堡

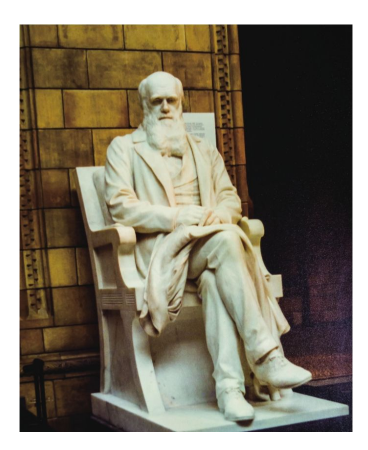
德（Alexander von Humboldt）等著名科学家，与他们进行了深入的交流。1827年，在古生物学家吉迪恩·曼特尔（Giden　Mantell）的推荐下，他拜读了拉马克的《动物学哲学》。虽然他此时还没有形成完整的进化思想，对拉马克的进化思想也未必认同，但是拉马克的进化思想对于赖尔渐变论思想的形成与完善还是产生了潜移默化的影响。赖尔在《地质学原理》的写作过程中，逐步表达出将地质现象归结于自然本身“水”和“火”的共同作用，以及地球在发展过程中是渐变的思想。这一观点的抛出在当时引来了极大的争议和不满。1829年，赖尔在伦敦地质学会上宣读了自己与他人合作的论文《以法国中部火山说明河谷的冲蚀现象》，巴克兰对其进行了激烈的反驳，师生之间闹得非常不愉快。科学上的争论与观点的捍卫，并不存在学生一定要服从老师的道理，如同亚里士多德所说：“吾爱吾师，吾更爱真理。”

1830—1833年这4年间，赖尔出版了《地质学原理》的前三卷。1837年，他出版了《地质学原理》第四卷，向灾变论发起了最后挑战。赖尔认为灾变论的最大问题在于：它将时间维度缩短了，将几百万年的发展时间误以为只有几百年……除了承认自然界中存在一次重大的灾变之外，它不包含任何有价值的理论。赖尔认为，人类是由其他生物进化而来的，地球在进行着持续不断的缓慢变化。赖尔对于地质学的分析和研究，对研究新生代地层的发展以及人类的起源和发展有着重要的理论意义。

## 2.3　自由自在的达尔文

提到进化论的诞生，有一位重要的人物永远无法绕过，这就是进化论之父——达尔文（Darwin）。达尔文1809年出生在英格兰中部地区什鲁斯伯里的一个富裕的中上阶层家中，家境比较富裕。他的祖父是一名博物学家，父亲是一名事业有成的医生，母亲也受到过良好的教育。在这样的环境下，达尔文却没有成为一个我们口中所说的乖孩子，整天忙着爬树、抓鸟、捕捉昆虫，以至于父亲经常大声地呵斥他：“你整天不是打猎、养狗，就是抓老鼠，你这样会让自己和全家人都丢脸的。”但是达尔文依然没有按照父亲的要求选择自己未来发

展方向的意思，他感觉射击和打猎，以及在大自然中自由自在地玩耍才是一件非常幸福的事情。



达尔文雕像

1825年，父亲把达尔文和他的哥哥（Erasmus）一起送到爱丁堡学医。能够与哥哥同行，他显然是乐意的。但是他很快就发现，自己非常厌恶学习医学。很快，两年时间过去了，1827年,达尔文并没有按计划拿到学位，就离开了爱丁堡。他多次对家人强调他一点儿也不爱学习医学，也不想和尸体打交道，他既不喜欢学医，也不喜欢父亲给他安排的当牧师的职业。为了摆脱令他讨厌的医学，他选择了相对来说不是特别讨厌的牧师职业，接受父亲的意见成为一名乡村牧师。但是在爱丁堡，达尔文也不是一无所获，当时他结识了一位年轻的动物学家罗伯特·格兰特（Robert Grant），是他给了达尔文启发，让达尔文了解到进化论先驱——拉马克的理论，引导他了解了最初的“物种变异”的进化思想。

从爱丁堡离开之后，达尔文开始在剑桥大学就读。在剑桥，达尔文有幸结识了剑桥大学的植物学教授约翰·史蒂文斯·亨斯洛（John Stevens Henslow）和地质学教授亚当·塞奇威克（Adam Sedgwick）。在他们的指导下，他了解了丰富的动植物学和地质学知识，而这些也是今后他在远洋航行中的重要武器！

1831年的春天，达尔文顺利通过了自己的学士学位考试，他虽然

顺利毕业了，但是心里对于未来的路究竟该怎么走还是非常忐忑。他不想毕业后回到家乡过上一眼就能看到老的日子，他想去实现自己旅行和探险的梦想，就像他心心念念渴望去加那利群岛探险一样。

一个非常偶然的机会，达尔文收获了他人生中的第一个“大礼包”。

当时，英国皇家海军“比格尔”号要进行一次远洋勘测和探险航行，希望亨斯洛推荐一个合适的人选。虽然亨洛斯自己觉得这是个好机会，但是想到要离开自己的妻儿，他心里又有点儿割舍不下。于是，他推荐了刚从剑桥毕业的学生杰宁斯（Jennings）。然而在临行前，杰宁斯由于种种原因突然变卦了。这时候，亨斯洛想到了达尔文，希望达尔文可以临阵补缺，以博物学家的身份进入皇家海军“比格尔”号远航。

达尔文的父亲一开始并不同意儿子进行远洋航行。一方面，达尔文随军舰出海需要他来买单，而且出海航行的时间是一个未知数；另一方面，他担心这样的旅行会让达尔文找到另外一个借口，耽误他更多的时间，会完全改变他未来的人生轨迹！但是，达尔文去意已决。面对年轻且倔强的儿子，父亲最终妥协了。

就这样，达尔文成功地踏上了旅途。他兴奋地宣称：“南美洲的甲虫要倒霉了！”

## 2.4 进化思想的萌芽

“比格尔”号停靠的第一站，是西非海岸佛得角群岛上的圣地亚哥。靠岸后，达尔文迫不及待地上岸去找寻动植物标本。每到一个地方，达尔文都会仔细地进行实地搜寻。“比格尔”号有时候会在一个海岛待上两天，有时候会在一个海岛待上十天半个月，达尔文正好也可以趁机缓解一下长期海上颠簸带来的晕船呕吐感。

3年后，他们到达了智利。在这里，他看到了地震的强大威力，也逐渐相信，大自然的力量完全可以将原先存在的物种都彻底毁灭了。




达尔文的地雀

在工作中，达尔文搜集了大量的实物资料。在别人休息时，他开始阅读拉马克和赖尔的著作，先驱们的物种进化思想逐渐在他的身上萌芽、生长。达尔文开始尝试利用自己搜集到的物证去验证这些思想，同时他也开始思考，是否可以利用手头的资料建立起全新的进化理论呢？当时教会宣扬的神创论漏洞百出，却从未改变，事实就是推翻它的最好武器，而达尔文已经做好了战斗前的准备。神创论认为每一个物种都是由上帝亲自创造出来的。达尔文在厄瓜多尔西岸的加拉帕戈斯群岛发现了大量的海龟和地雀，而这些海龟和地雀之间都存在着或多或少的差异。例如，各个岛屿上的地雀在体形、颜色、食性、鸟喙上都有着各自的特点。这是神创论无论如何也解释不了的——上帝怎么会有时间不厌其烦地创造出这么多各有特色而又属于同一种类的生物呢？唯一合理的解释就是生物是逐步进化而来的！对达尔文产生深刻影响的还有各种自然形态的变化。例如，他在智利安第斯山海拔3657米处发现了大量海蛤类动物的化石，这便证明了现在的山顶曾经是海底，地形是在逐步变化的，经历了沧海桑田的变迁。同时，这也印证了赖尔地质渐变学说的正确性。通过这些化石，达尔文对神创论充满了质疑与不屑，更加坚信物种进化的观点。科考回来后，达尔文开始着手写作。他将自己关于物种进化的观点和在考察途中搜集的物证资料结合在一起，用事实来论证自己的理论。

1859年11月24日，划时代巨著《物种起源》出版了，他用大量翔实的证据论证了生物在不断进化、物种是渐变的观点。达尔文认为，自然界可以在相对较长的时间里，通过自然选择挑选出与自然环境相适应的物种。换句话讲，就是“物竞天择，适者生存”。实际上，进化论的提出应该是两位科学家共同的贡献，这一理论是由两位科学家分别独立提出的，除了达尔文以外，还有一位叫华莱士（Wallace），他是英国的博物学家、探险家、地理学家和人类学家。他在《物种起源》出版的前一年——1858年，曾给达尔文寄去一篇论文《论变种无限地离开其原始模式的倾向》。在这篇论文中，华莱士详细地阐述了物种进化和自然选择的原理，可以说华莱士已经先于达尔文系统地提出了进化论的雏形。华莱士的经历和达尔文有着诸多相似的地方，他曾经在马来半岛和印度尼西亚群岛考察过。在考察过程中，他对大量的化石证据和物种形态学方面的证据进行研究，发现物种是逐渐进化的这一事实。在拉马克和赖尔进化思想以及马尔萨斯（Malthus）《人口论》的影响下，华莱士独立地提出了一整套的进化理论。这是历史上第一套完整的进化理论，在他完成整篇论文的时候，达尔文的著作尚未完成。

## 2.5　解释不了的难题

经过达尔文、海克尔（Haeckel）、赫胥黎（Huxley）、华莱士等的不懈努力，在《物种起源》出版后的几十年里，进化思想在社会上得到了普遍认可，人们开始逐步抛弃神创论、灾变论，转而接受进化理论。伴随着这一理论逐步深入人心，研究它的人逐渐增多，于是又出现了不少质疑的声音。

第一个让达尔文头疼的问题就是被誉为“热力学之父”的物理学家威廉·汤姆孙（William Thomson）提出的。汤姆孙从热力学角度进行测算，得出地球的年龄只有几千万年。这样的时间长度，对于进化论来说，无疑是白驹过隙，大自然几乎不可能在这么短的时间内完成物种的自然选择。这种质疑无疑是釜底抽薪式的，达尔文没有办法从专业的角度进行反驳。但是令人感到戏剧性的是，这件事却并不怪达尔

文，是汤姆孙在计算的时候忽略了地球内部的热量，因此计算出来的地球年龄远远小于实际的地球年龄，但是面对热力学的权威质疑，达尔文依旧束手无策。



地球的演化

第二个质疑来自爱丁堡大学的工程学教授弗莱明·詹金（Heeming Jenkin）。他针对《物种起源》中的遗传问题进行了有针对性的批评，在《物种起源》出版近十年的时间里，他写了一篇书评，这篇书评提出了一个问题，让达尔文陷入了尴尬的境地。他假设“一个白人因为海难流落到一个黑人居住的岛上，……不能就此得出结论，经过很多代或者无数代之后，岛上的居民就会都变成白人。”在大英帝国四处扩张的鼎盛时期，詹金认为，一个白人单枪匹马就可以征服岛上所有的黑人居民，先杀掉一部分，再娶上一群妻妾，然后给一大群孩子当爹。然而，即便有白种人那种明显的“优越性”，他也无法想象这就能将岛上后来的居民都变成白种人。达尔文意识到詹金的观点有一定的道理。

事实上，如果我们说遗传的作用好比把牛奶倒进咖啡里，两组性状通过遗传混合在一起，就会导致两者的稀释，这种稀释的速度是非常快的，这就从根本上否定了大自然日积月累漫长的选择概念。

所以达尔文不得不采取错误的“泛生论”来弥补自己观点中的漏

洞。他认为，人体的每一个细胞中都含有携带了遗传物质的微小颗粒，我们把它称为“胚芽”。胚芽会从父母身上转移到后代身上，直到找到属于自己的地方。这种观点明显也是错误的，但是达尔文在当时不得不用这样的理论来弥补他学说中的漏洞。



咖啡与牛奶的混合

但是随着时间的推移，19世纪中叶达尔文提出的进化论，又受到了来自多方的挑战。

首先，对达尔文进化论进行完善的是综合进化论。它从三个方面对之前不完善的地方进行了系统的修改。综合进化论最大的特点，是融合了格雷戈尔·孟德尔（Gregor Johann Mendel）的遗传学理论，解释了达尔文时代不能解释的可以遗传或者不可以遗传的问题。它从基因的角度，深入地解释了为什么有的性状可以遗传给子代，但是有些性状就无法遗传。它还对达尔文常常使用的错误观念——获得性遗传表示了批判。综合进化论还完善了之前进化论的不足。例如，达尔文认为个体是进化的主体，但是综合进化论就认为种群才是进化的主体。个体的数量太小，不能保证可以将性状稳定地遗传下去，但是种群的大数量就可以起到稳定遗传的作用。



斑马个体



斑马种群

1968年，日本生物学家木村资生提出了进化的中性理论，这对进化论来说是一项近乎颠覆性的挑战。他依据核苷酸和氨基酸的置换速度，提出了分子进化的中性选择学说：多数或者绝大多数突变都是中性的，无所谓有利或者不利，因此这些中性突变不会发生自然选择和适者生存的情况。生物的进化主要是中性突变在自然群体中进行着随机的“遗传漂变”的结果，而与选择无关。这一学说的提出对达尔文的进化论来说无疑是一次颠覆性的冲击。客观上说，由于生物的进化来源于突变，而突变很多是中性的，所以用“生物进化”这个词显得就不那么确切了，应该用“生物演化”可能会更加准确一些！

除了中性突变学说的挑战之外，来自化石方面的证据也逐步对进化论物种渐变的思想提出了挑战。按照进化论的说法，经过漫长的历

史演变，各个时期的动植物演变过程都能够在不同时期的岩石地层中找到对应的化石证据。但是令人费解的是，化石中的链条却大多是缺失的。最典型的例子来自始祖鸟，我们可以看出始祖鸟既有鸟类的特征，又有爬行动物的特征。这个事实可以用来佐证鸟类也许是来自爬行动物，但是在始祖鸟和爬行动物之间以及始祖鸟与鸟类之间，并未发现任何中间形态的生物化石存在，这让坚定的渐变论者似乎从心底开始动摇了，这些问题从化石的角度是无法得到完美解释的。那么物种究竟会不会有跳跃式的发展变化呢？



德国巴伐利亚侏罗纪平版矿床始祖鸟化石

与此同时，现实中有很多能够佐证物种可以发生跳跃式变化的例子。从物种的数量上来看，现存的物种只有原先物种总数的十万分之一到千分之一，绝大多数的物种已经灭亡了，例如我们耳熟能详的恐龙的灭绝，在二叠纪的一次物种大灭绝中，有超过半数的物种灭亡了。因此物种的灭绝可以看成对渐变论的一种有力的驳斥。这种灭亡的状态完全是突变形式的，类似于之前居维叶的灾变论，没有任何的铺垫就突然发生了。

迄今为止，关于进化论的争论依然在进行，进化的理论也在逐步地发展和完善中。科学发展的历程中没有任何一种理论可以做到毫无瑕疵，都是在不断的质疑和驳斥中发展完善，这也许就是我们逐渐进

步的阶梯，人类也只有在质疑中才能不断地砥砺前行！

# 3　6600万年前的绝唱——霸主的灭绝

在聊完物种的起源和演化的过程之后，我们肯定会有很多的疑惑，既然从物种诞生以来，地球上出现了大量的物种，但是为什么现存的物种数量只有所有物种总数的十万分之一到千分之一呢？为什么会有那么多种物种，包括我们熟知的恐龙都灭绝了呢？



霸王龙

恐龙作为一种已经灭绝了的史前动物，它们让我们感到既熟悉而又陌生，熟悉的是在各种经典的电影和电视剧、畅销的书籍、孩子们的玩具中都能见到它的身影；陌生的是恐龙离我们现实的生活非常遥远，毕竟它们生活在距今几亿年前的远古时代，我们从未真实地感受过恐龙的存在，它们的突然消亡也给我们留下了无限的遐想，这也正是其中的魅力所在。

## 3.1　霸主登场

伴随着整个世界从大灭绝中逐渐复苏过来，2.3亿年前恐龙逐渐登上历史的舞台。古生物学家史蒂夫·布鲁萨特（Steve　Brusatte）推测，最早的恐龙实际上出现在2.4亿年前至2.3亿年前。史蒂夫·布鲁萨特是电影《侏罗纪世界3》的科学顾问，他命名了15个新物种，对恐龙这个物种有着独特的研究。

虽然恐龙的这次登场并不是那么耀眼，但是这毕竟意味着在接下来漫长的中生代，注定是属于恐龙的天下！而真正的恐龙究竟是从哪种原始的物种类型中脱离出来的，科学界至今都没有定论，到目前为止依旧是一个未解之谜。

在恐龙初次登场的时候，例如埃雷拉龙和始盗龙出没的年代，也就是距今2.4亿年到2.3亿年前，当时的地球和现在的地球是截然不同的两个世界，当时整个地球上的大陆是连在一起的，就像一个巨大的字母C，地质学家又将它称为“超大陆”。这一整块大陆上的气候就像桑拿房一样，二氧化碳含量高，温室效应明显，气温比现在高得多。

地球上最适合物种生存的地方，就是南北极，那里气候温暖潮湿，就像现在的昆明一样四季如春。但是，伴随着地质板块的不断运动，火山喷发，大陆在不断变化，气候也在不断改变，在侏罗纪晚期，整个大陆的板块已经较三叠纪有很大的变化，彼此之间也在不断地分离，到了晚白垩纪，已经基本上有了现在地球上七大洲的雏形了。伴随着大陆板块的运动，气候也在逐步发生着变化，地球的温度在逐步地降低，不再是完全一片的桑拿房天气，南、北极也开始由温带气候向寒带气候转变，而其他大陆也在发生着演变。在这样一片广袤的土地上，原始的恐龙由跟猫差不多大的恐龙祖先开始逐步地演变而来。

恐龙的演化呈现出很强的地域性，挡住它们自由扩散的，不是地理上的屏障，而是难以忍受的气候，原先它们更多地集中在超大陆南部的某一个小的区域。毕竟恐龙始终喜欢生活在那些气候潮湿的地方。但是随着气候的逐渐变化，它们已经开始在地球的各个角落里自由发展了！

下面让我们一起还原恐龙的登场过程。在古生代的二叠纪，恐龙已经出现了，但是它们依旧生活在哺乳动物和鳄鱼亲戚占据统治地位的时代。这时候的恐龙还只是其中一种非常小众的生物，也算不上很凶猛，常常被其他动物欺负。当时地球上的霸主是一种叫作蜥鳄的生物，蜥鳄是一种凶残的、25英尺（1英尺约合0.3048米）长的鳄系主龙类动物，它们面对恐龙有着绝对的优势。之前我们提到了物种“界、门、纲、目、科、属、种”的分类方式，主龙类分别向两个方

向发展：一支分化成鸟跖类，另一支分化成假鳄类。其中，假鳄类始终占据在生态位的顶部，并且逐步演化成鳄类，鸟跖类逐步演化成恐龙。在古生代的三叠纪中，始终是假鳄类站在食物链的最顶端，恐龙在它们面前显得微不足道，根本没法和它们抗衡。当时还有一种非常可怕的两栖类生物叫作宽额螈，作为青蛙、蟾蜍、水螈和蝾螈的祖先，它的样貌丑陋、性格凶残，头约有1平方米这么大，两颌里长了数百颗牙齿，又宽又大，近乎扁平的上下颌，可以吞下几乎所有想要吃的东西，那个时代的恐龙都希望离这种恐怖的动物越远越好，遇到了也得绕道走。




三叠纪时期的恐龙

所以面对这么多强大的对手，恐龙只能选择隐忍下来，慢慢积蓄力量，盼望着属于它们的时代到来。

## 3.2　走上巅峰

伴随着三叠纪的结束，恐龙和它最大的对手——假鳄类一起进入了我们最熟悉的侏罗纪，那么恐龙是如何从凶残的假鳄类的统治下脱颖而出，取得统治权的呢？这也是恐龙演化最关键的一步。

恐龙是如何做到这一点的，古生物学界有一些大胆的推测。目前，相对来说得到大家公认的说法是因为三叠纪末期的火山大喷发。由于地球的板块运动，泛大陆一直在不断地裂开，岩浆从地底喷涌而

出。据古生物学家推测，有些地方喷发出来的岩浆堆积起来有3000英尺厚，足以吞噬两个帝国大厦那么高的建筑；泛大陆中央约300万平方千米的区域直接被熔岩所淹没；地表覆盖着熔岩的同时，大气中也充斥着各种喷发带来的有毒气体，全球温度急剧变化，变得更加温暖，这导致约3成的物种彻底灭绝了……

这种情景与1亿年之后恐龙灭绝的情况如出一辙，世事轮回。如果没有这次大的灭绝事件的出现，也许恐龙永远无法摆脱假鳄类的统治，更无法在与假鳄类的对抗中占据优势，也不可能在后续的生存发展中体型越来越大，分支越来越多，直至统治整个地球。恐龙可以说是这次大灭绝事件的最终获利者，原先的统治者——假鳄类幸存下来的仅仅只有几种原始的鳄鱼，最终演化成了我们现在所看到的短吻鳄和鳄鱼，其他的各种分支类型都销声匿迹了。

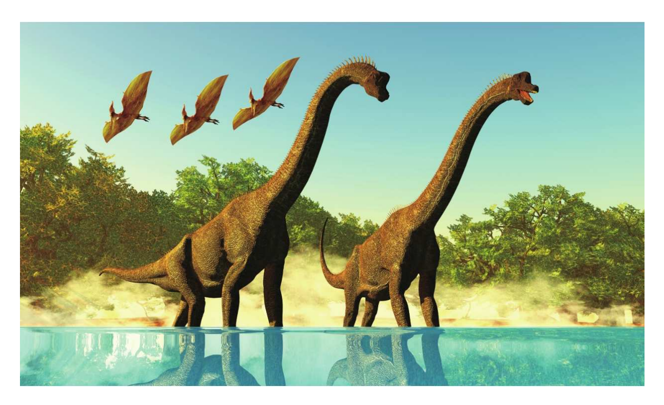


假鳄类演化出的短吻鳄

从二叠纪的末期逐步登上历史舞台，直到侏罗纪，地球才真正成为恐龙的天下，它们主宰着整个地球直到白垩纪末期。恐龙在自己的乐园上不断地发展演化，出现了很多新的类型：小盾龙、双嵴龙、剑龙、雷龙、腕龙、梁龙……在侏罗纪后期，雷龙、腕龙这样的巨兽已经可以生长到30多吨，甚至更大。在白垩纪曾经出现一种巨龙类的种属，包括无畏龙、巴塔哥尼亚龙、阿根廷龙等，它们的体型庞大，体重超过50吨，有的比我们现在的波音737客机还要大，这么大的物种自恐龙灭绝之后就再也没有在地球上出现过，我们也很难想象，有什么物种能跟这样的庞然大物相抗衡？又有什么力量能将它们彻底灭绝呢？



雷龙



腕龙

在巅峰时期的侏罗纪，恐龙种群中出现了几种特别凶猛的分支。其中有一种号称“侏罗纪屠夫”的异特龙。它们的成年个体体重在2吨到2吨半之间，身长30英尺左右，善于奔跑。之所以获得了“屠夫”这样的称号，是因为它们可以将自己的脑袋当成手斧砍杀猎物，它们的牙齿并不粗壮，但是头骨却非常坚硬，可以承受巨大的冲击。白垩纪还出现了另外一种鲨齿龙类，主要包括三个物种：南方巨兽龙、马普龙和魁纣龙。鲨齿龙类演化得更加强壮和凶猛，出现了鲨鱼一样的牙齿，可以轻松地撕碎食物，轻易地走上了食物链的顶端，成为那个时

代的超级肉食者。另外，还有一种同样出现在白垩纪的暴龙，也是整个恐龙发展史上的终极统治者。美国古生物学家费尔菲尔德·奥斯本（Henry Fairfield Osborn）在1905年向全世界宣布他发现了一种新的恐龙，并将其命名为“君王暴龙”，意思是“凶暴的蜥蜴之王”。2009年，在我国东北部地区也出土了一种类似的恐龙化石，古生物学家将其命名为中国暴龙。暴龙从一登场就站在了食物链的顶端，并且盘踞于此长达上亿年之久。



暴龙

为什么暴龙有这么强大的攻击力呢？我们拿最具有代表性的君王暴龙做一个简单的展示。君王暴龙主要生活在6800万年前至6600万年前，它们主宰着北美中西部覆盖着森林的海岸和河谷。成年的君王暴龙长达42英尺，体重可以达到惊人的7～8吨，这虽然和植食性的恐龙没法相比，但绝对称得上是陆地上最大的纯肉食性动物，这在地球的发展史上也是绝无仅有的。君王暴龙的头部有一个微型汽车那么大，从口鼻部的前端到耳朵长约5英尺，嘴巴里有50多颗锋利的牙齿，口鼻部前端长着用于啃咬的短齿，在上下颌两侧各有一排锯齿形的“道钉”，大小跟我们常吃的香蕉差不多。这种牙齿看起来就像一排排尖锐的刀子，它们的手指和脚趾就像一把把弯曲的钢钩，体表覆盖着一层厚厚的鳞片。这些特征都预示着，这一定是一种恐怖的肉食性动物。

我们可以在电影中或者书中看到这样凶残的进食画面：暴龙叼起

自己的猎物，不断地摆动头部，疯狂地撕咬，空中血花四溅，还伴随着牙齿的掉落。君王暴龙具有人类不具备的本领——终生不停地更换牙齿，可以补充在剧烈捕食中断裂或者损失的牙齿。君王暴龙的进餐方式也与众不同，它们采用的是“穿刺—拉扯式”的进食方式。在“穿刺”阶段，君王暴龙会用上下颌巨大的咬合力将猎物的骨头咬碎。这种咬合力究竟有多强呢？我们举一个简单的例子进行比较。人类臼齿的咬合力在175磅（1磅约合0.4535千克）左右，非洲狮的咬合力在940磅左右，而君王暴龙的咬合力高达3000磅，是非洲狮咬合力的3倍多。“穿刺”之后进入的是“拉扯”阶段，君王暴龙会来回猛烈地撕咬，将猎物直接撕碎后吞进腹中，甚至连带着大块骨头，这也是为什么在很多恐龙粪便的化石中可以找寻到大量骨头碎片。而支持这些动作的最关键之处是拥有巨大的颌部肌肉，可以配合钉耙似的牙齿和坚硬的头骨轻松地撕碎和咀嚼猎物。



暴龙短小的上肢

除了上述特殊的身体结构外，君王暴龙还有一个让我们非常好奇的身体部位，那就是它们的上肢。也许我们在电影中会嘲笑君王暴龙短小的上肢，因为这种短小的手臂看起来非常的可笑。为什么它们不会在进化的过程中将其彻底退化掉呢？就像鲸鱼一样，当它们逐渐从陆地走向水域的时候，就抛弃了无用而又累赘的后腿。原来这是有原因的，君王暴龙看似非常短小可怜的上臂，其实有着强大的肩伸肌和肘屈肌，可以用来压制住不断挣扎的猎物，好给上下颌咬碎猎物的头

骨提供充足的固定时机，因此上臂也是它们称霸地球的一种有效的武器！

## 3.3　飞向蓝天

在掌控着整个地球的同时，恐龙的活动范围也逐步地扩大，它们不再仅仅局限于在陆地和海洋活动，也逐步开始向蓝天进军！

其实，关于恐龙不断演化以至于最终飞向蓝天，有一种普遍的说法：海鸥还有其他所有的鸟类都是从恐龙演化而来的，换句话说就是鸟类和恐龙有着共同的祖先。史蒂夫·布鲁萨特有一个非常重要的观点就是，鸟类就是恐龙。那么是不是就意味着，恐龙从某种意义上说并没有完全灭绝呢？所以，时至今日，“恐龙”也依旧生活在我们的周围。鸟类是恐龙主宰世界1.5亿年之后留下的永远不能被磨灭的遗产，是一个已经逝去的古老帝国的遗老。因此，鸟类是恐龙已经成为恐龙考古学家最为重要的一种发现。

鸟类是由恐龙进化而来的观点其实并不新鲜，早在达尔文的时代就已经提到一种重要的化石——始祖鸟化石，始祖鸟距今已有1.5亿年的历史。恐龙时代生活着很多鸟类，在侏罗纪中期的某一时期，一种体型小、有翅膀、能鼓翼飞行的真正的鸟类已经演化出来了。

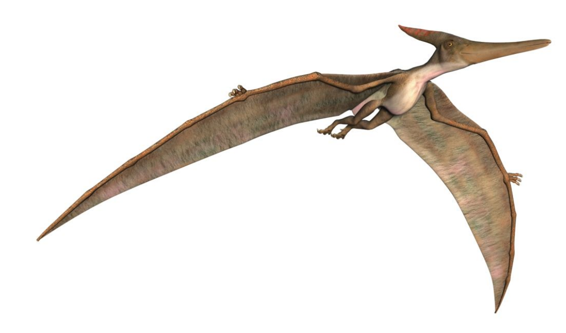


会飞的恐龙

那么我们能不能简单地还原一下，鸟类是如何从恐龙中演化出来

的？鸟类又是如何形成了它们独特的身体结构呢？例如，鸟类拥有其他生物所没有的羽毛、翅膀、没有牙齿的喙、叉骨、中空的骨头、牙签状的细腿……这样的体型也让鸟类拥有了飞翔的能力、极快的生长速度、温血的生理功能、高度发达的小脑、强健的胸部肌肉……

在我国辽宁出土的恐龙化石似乎给了我们一个可能的答案，某些在陆地上直立行走的恐龙的左右锁骨紧密地融合在了一起，形成了一个新的结构——叉骨，当时这个细微的改变主要是为了让在陆地上行走的恐龙在捕获猎物的时候能够更好地吸收冲击力，在生物的演化进程中，叉骨逐步在鸟类飞翔时演化成储存能量的弹簧。之后，胸部有叉骨的手盗龙又演化出了弧线形的脖子，据科学家的推测这可能与搜寻猎物有关，而且个头也变得更小，这样更有利于它们占据新的生态位——树木、灌木丛，可以搜寻到更多的食物。再之后，这些小型的、直立行走的、有着叉骨脖子摇来摇去的小体型的兽脚类恐龙开始将胳膊折叠在身体的两侧，逐步形成了现在鸟类原始的模样。



会飞的翼龙

鸟类还有一项重要的特征，那就是羽毛。羽毛并不是在鸟类登上历史舞台时才出现的，而是在恐龙祖先身上就已经存在。当初羽毛的主要功能是用来保持体温和伪装身体，但是后来在鸟类身上却逐步进化出来用于飞翔的功能。翅膀不仅仅可以用来滑翔，类鸟的恐龙甚至开始利用羽毛进行飞翔。古生物学研究中的数学模型也显示，一些非鸟类的恐龙也能够通过拍打翅膀飞翔，包括小盗龙和近鸟龙。

有了上述身体结构上的改变，恐龙飞向蓝天也有了最基本的保

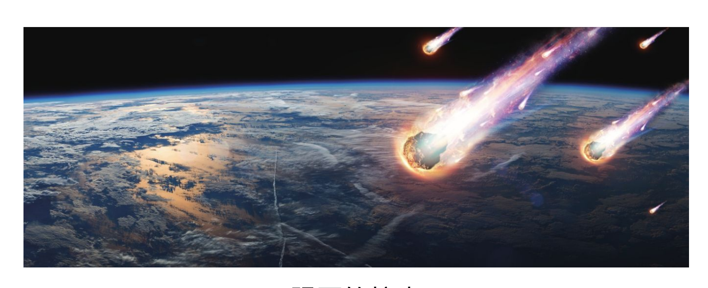
证。演化的过程中出现了这样一类恐龙，它们身材小巧、翅膀上覆盖满了羽毛、生长迅速、呼吸频率极高。它们不断地在地面跳跃，不断地腾空飞上树梢，此时空中的飞行并不是它们生活的全部，只是在树上也存在着它们的一些生活的区域而已。但是这种行为，为自然选择开启了一扇飞向天空的大门，让它们具备了征服蓝天的能力和潜质！

## 3.4 霸主的谢幕

在距今6600万年前的地球，整个恐龙家族已经傲立在这颗星球长达上亿年之久，谁也想不到会出现历史性的改变，让这些称霸世界的霸主，这些站在整个食物链顶端的生物会在短短的时间里销声匿迹。

关于恐龙灭绝的说法很多，很多研究者给出了多种可能的猜测，但是这些说法都无法得以证实，其中有一种说法流传得最为广泛，那就是来自外太空的一颗行星的撞击，终结了恐龙在地球上的霸主地位，让它们从地球上彻底地消失，之前的所有传奇都归于平淡。

有这样一种说法，白垩纪的一天，在北美洲的西部，一颗彗星或一颗小行星与地球相撞，撞击地点位于今天墨西哥尤卡坦半岛所在区域，至今这里都存留着当时撞击的痕迹。这颗彗星的直径约6英里（1英里约合1.6093千米），大小跟珠穆朗玛峰差不多，移动速度在每小时67000英里左右，比喷气式飞机快100倍，相当于10亿颗原子弹爆炸时所产生的能量。地壳被撕开了一道长约25英里的口子，直达地幔层，同时形成了一个直径超过100英里的陨石坑。如今我们也已经在很多地质板块的岩层中发现了端倪，例如，在墨西哥发现的陨石坑，直径大约有180千米，深900米，被掩埋在数百米的沉积岩下面。



陨石的撞击

撞击发生之后，首先出现的是地球外壳蒸发、熔化或者被弹射出去，形成了一个大火球，横扫整个世界。随后这种巨大的火球不断地引发地壳震动，引发大火，撞击带来的能量需要从岩石和土壤中释放出去，整个大地都变成了蹦蹦床的模样，面对这样的灾害，地球上的绝大多数生物都是无法幸免的。处于霸主地位的恐龙并不清楚究竟发生了什么，只能四散逃命。所有的恐龙无论体型多大，在大自然面前都是如此的渺小，被大地抛来抛去，而且这种硕大的体型也成为它们灭绝的重要原因。

很多古生物学家认为，这次大型的撞击事件仅仅是一个导火索，伴随着撞击，大地不断地震颤，继发而来的是地壳的变化，火山喷薄而出，海啸随之而来，而撞击带来的后遗症也在逐步地显现。也许在当时有一些恐龙在撞击中侥幸存活了下来，在撞击发生之后，很多撞击点之外的恐龙也逐步感受到了连锁反应的可怕。在接下来的一个月、两个月、一年、两年的时间里，地球变得又冷又暗。撞击和火山喷发带来的各种烟灰和岩石的粉末依旧滞留在大气中，遮蔽了阳光，随之而来的就是寒冷，就像是“核冬天”，只有非常顽强的生物才能够生存下来。漫长的黑暗时间让很多植物不能够进行光合作用，植物大量地死亡，紧接着带来的就是植食性动物的死亡。多米诺骨牌一张张依次倒下，直到当时地球的统治者——恐龙，也感觉到了绝望。



恐龙的灭绝

浩劫过后，仅仅只有部分体型较小的物种存活了下来，比例大概只有原先物种的30%，甚至更少。活下来的哺乳动物比死去的哺乳动物个头更小，食谱也更加的广泛，它们可以四处游荡，同时又能够在洞穴中躲藏。而且生活在水中的生物比生活在陆地上的生物有着更强

的生存适应能力。撞击发生之初，它们可以借助海水、湖水缓冲岩石的撞击和地震的袭击，之后，水中的生态系统受到自然环境的影响要滞后很多，一些水生植物死亡之后，腐败的尸体也能够给水生生物提供食物……

而这些优势，在恐龙的身上都荡然无存。恐龙硕大的体型在面对陨石撞击、岩石飞屑、火山喷发的时候，很难寻找到合适的掩护，直接暴露在打击之下，瞬间遭受到毁灭。此外，恐龙的演化让它们的食谱变得更加专一，要么只吃肉，要么只吃几种特定的食物，这种情况就不如幸存下来的哺乳动物，它们有娇小的身躯，有合适的庇护所，有广泛的食谱……

最后，恐龙灭绝还有一个容易被大家忽略的因素：大多数恐龙都是温血性动物，或者至少如此。因此它们的代谢水平高，需要大量的食物，不可能连续不吃不喝蛰伏几个月之久，而这一点两栖动物和爬行动物就可以做到。另外，恐龙生蛋和孵化需要3~6个月的时间，这差不多是鸟类孵化时间的2倍，孵化出来之后，小恐龙需要更长的时间才能够长到成体，漫长的生长和发育期也让恐龙在如此严峻的自然灾害面前变得极其脆弱、不堪一击。

也许不是一种灾难，是多种综合的因素让恐龙的灭绝变得顺理成章，而人类的祖先也逐步地从老鼠般的大小成长起来，顺利地从白垩纪进入了古近纪。与此同时，有一部分的鸟类存活了下来，绝大多数鸟类和一些长着羽毛的、鸟类近亲的恐龙已经死亡了，包括所有长有四翼的、像蝙蝠一样的恐龙。还有那些拖着长长尾巴、长着牙齿的原始鸟类也跟随着白垩纪的行星撞击而一同烟消云散。但是与现代鸟类相仿的鸟类存活了下来，这究竟是什么原因，目前尚无定论，但是有一种可能的说法是：宽大的翅膀和强健的胸肌让它们得以快速地逃离那些危险的场所。

那么究竟是这次撞击一下就导致了恐龙的全部灭绝，还是有很多更为深层次的原因呢？为什么有的物种能够躲避这次大的灾难？而有的物种则全部销声匿迹了呢？在地球上，我们已经找到了陨石撞击的痕迹，也证实了这些撞击导致了将近7成的物种灭绝，包括大量的哺乳类动物和爬行类动物，而恐龙则在这次撞击中，从巅峰一下子跌落

神坛。目前存在着太多不确定的谜题，而对这些问题不断探索的过程，终究会让我们一步步逼近真相，让我们逐渐了解史前那个古老而又神秘的年代！

# 4　控制寿命的时钟——端粒

大家在电影、电视剧中看到过很多长生不老的神仙，也在现实生活中看到过很多仙风道骨的白胡子老爷爷，大家心里可能会不由得产生疑问：人类的寿命究竟有多长？人类究竟能不能实现长生不老的梦想？

## 4.1　衰老的谜团

在日常的生活中，大家会注意到一种奇特的现象，有些人虽然年纪很大，但是却显得很年轻，有的人岁数很小，但是却显得老态龙钟。正常情况下，我们的面容和我们的年龄基本上是相符的，即使有些时候出现了一些所谓的“逆生长”的事实，这与心态的年轻、保健方式的多样、化妆手段的提升都有密不可分的关系。从本质上来说，年龄和衰老保持对应的关系依然是符合自然规律的。



老年斑

从呱呱坠地的孩童到耄耋之年的老者，我们的容颜会逐渐发生变化，会增长些许皱纹，会有白发产生，身体的机能也会发生退化。有的老人还会出现驼背、行动迟缓、言语含糊、记忆力衰退等症状。

衰老是每个人都要经历的必然的过程，在衰老外在的表现中，有

一个重要的特点，那就是老年斑的产生。大家可以在老年人的体表看到一块块或一点点黑褐色的沉积，这就是老年斑。它的本质是一种色素的沉积，尤其是在脸庞和手上。

大家不要简单地认为老年斑只是在皮肤的表面产生，其实它在很多器官中都可以沉积，只不过内在的器官我们无法从体表上看到而已。可以这么说，当你的手脚或面部皮肤出现老年斑的时候，这就意味着你体内的器官也已经产生老年斑了。

老年斑的产生和身体机能的退化有直接联系，是身体不能排出或者清扫某些代谢产生的垃圾而导致的。

举个简单的例子，大家就能够明白老年斑的本质。每个人的身体就相当于一个个庭院，每天都有一位清洁工来打扫庭院。当清洁工很年轻的时候，这些垃圾都能够很轻易地被清除，所以我们的皮肤看起来很光洁。但是伴随着这个清洁工逐渐老去，她打扫庭院的能力越来越差，当有一天她无力再去清扫，或者清扫得不干净的时候，垃圾就会不断地在庭院中堆积起来，形成老年斑。老年斑的本质就是我们衰老的标志，说明体内清除垃圾的能力已经减弱了。

那么人类究竟能活到多大岁数呢？综合目前关于人类寿命的三种假说来看，人类的理论寿命应该在120～150岁，可是现实生活中能够活到这个岁数的人却寥寥无几，甚至人均寿命只有理论寿命极限的一半。其中的原因究竟有哪些呢？

我们所说的理论寿命是在各种条件都很完善，身体中的细胞没受到任何损伤的前提下才能够达到的。现实中，我们生活的环境中存在着各种各样的对身体有害的因素，例如，化学药品的损害、空气中的雾霾、烟酒的刺激等，都会对人体产生很大的危害。虽然说随着生活水平的提高、医疗条件的完善，人类的平均寿命在大幅提高，但是要达到理论上的极限寿命依旧任重而道远。

然而在现实生活中，还有这样一类人，他们被称为早衰症患者，又叫儿童早老症。他们可能只有几岁，但是看起来就像七八十岁的老人一样，满脸的皱纹，花白的头发，甚至有人还出现了驼背和蹒跚的现象。这究竟是为什么呢？

这些儿童都有一些共同的特点，例如，发育延迟、头发稀少、皮肤老化、头皮血管突出、骨质疏松等，这些都是老年人所具有的特点。

这些孩子尚处在含苞待放的豆蔻年华，为什么会出现衰老的现象呢？是不是身体中控制衰老的机关被无意中打开了呢？

研究表明，这和遗传存在着密切的联系。那些患上早老症的孩子一般很少有人能活过20岁，毕竟20岁对于五六岁就已经显现出衰老迹象的孩子来说已经是很大的年纪了。曾经有人做过比较，如果患上这种疾病，每过1年就相当于正常人过10年左右。就像《西游记》中描述的，天上一天，地下一年。早老症儿童的1天就相当于正常人的10天。

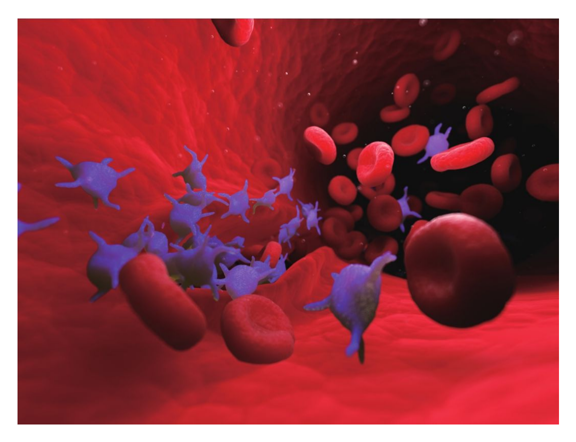
这种遗传疾病的发病率在800万分之一到400万分之一。对于家族里没有这方面疾病遗传史的人来说，除非发生了基因的变异，基本不用担心这种疾病会发生在自己和孩子的身上。

早老症的出现释放了一个重要的信号，那就是在身体中一定有控制着衰老的信号机关，当这一机关被触发的时候，就开启了衰老的进程。

## 4.2　细胞能活多久

要了解人类究竟能活多长时间，首先要了解一下组成身体的细胞究竟能活多久，不同的细胞寿命有着天壤之别。

以人体为例，当细胞在没有受到其他损伤的前提下，正常肝细胞的存活时间是几百天；血小板细胞的寿命是7～14天；肠黏膜细胞的寿命是3天左右；而血液中的中性白细胞可能昨天才出生。于是出现了一种奇特的现象，不同的人体细胞各自具有不同的细胞生命周期。有的细胞如味蕾细胞、表皮细胞不断死亡，不断更新换代；而有的细胞如神经细胞却和我们一起慢慢变老。



血小板（蓝色）



青蛙逐渐消失的尾巴

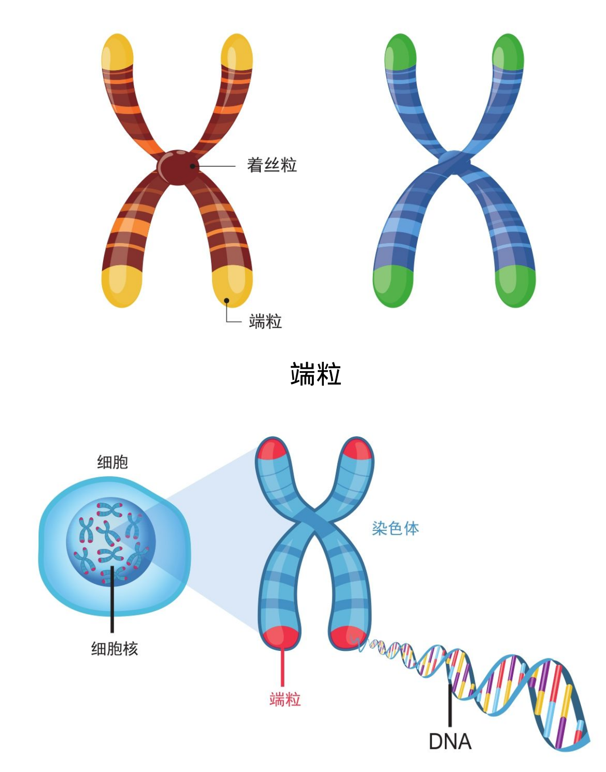
在日常生活中，有很多细胞衰老和死亡的例子，多到让我们忽视了这些现象的发生。例如，小蝌蚪在成长为青蛙的过程中，尾巴逐渐变短并消失，就是通过细胞的死亡实现的。还有人体的某些表皮细胞，如头皮屑、皮肤表皮角质等都属于死亡的细胞，这些物质持续产生，意味着不断地有新细胞出生、有老细胞死亡，间接地说明不同组织的细胞有着各自的寿命。

## 4.3 生命的时钟

既然人类是有寿命极限的，各种细胞也是有一定寿命的，那么在我们身体中有没有一种物质在背后悄悄掌控着这一切呢？

答案是肯定的，这种被称为生命时钟的结构就是——端粒。

端粒的准确定义是存在于真核细胞线状染色体末端的一小段DNA和蛋白质复合体。这样的定义，听起来有点复杂，简单来说，端粒就是染色体末端一小段重复的DNA片段，如果把DNA比作一条鞋带的话，端粒就是鞋带末端的塑料鞋带扣。它由简单的DNA高度重复序列组成，端粒结合蛋白一起构成了特殊的“帽子”结构，作用是保持染色体的完整性和控制细胞分裂周期。DNA分子每次分裂复制，端粒就缩短一点，一旦端粒消耗殆尽，细胞便会迈入程序性死亡的过程。



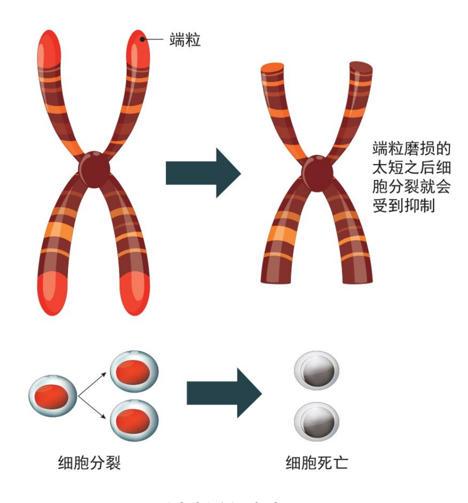

端粒

端粒在细胞中的位置

那么端粒最初是如何被发现的呢？1946年诺贝尔生理学或医学奖获得者赫尔曼·约瑟夫·穆勒（Hermann J. Muller）在研究染色体结构的时候，发现了一个奇怪的现象：断裂的染色体末端很容易发生相互之间的黏合，形成各种不同的染色体畸变，而天然的染色体结构却极

其稳定。这就说明正常的DNA序列和它末端的一段DNA在性质上有很大的差别，末端的DNA并不具有什么具体的功能，但是它能够起到稳定遗传物质DNA的作用。穆勒还发现，如果末端的DNA减少到一定程度，就会逐渐地失去稳定性，发生解体，细胞也就会走向死亡。

穆勒其实并不清楚，他无意中发现的这段特殊的结构——端粒，正是我们苦苦寻找的“生命时钟”。我们可以把端粒看成DNA的保护套，这个保护套起到了稳定遗传物质的作用，可以防止不同的染色体之间发生黏连，确保这些染色体结构的稳定。



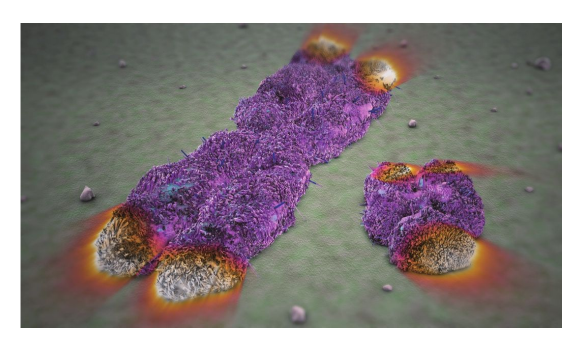

端粒的磨损

细胞每分裂一次，端粒就会出现不同程度的损伤，不同物种的端粒磨损程度是不相同的，由于DNA分子复制的机制，DNA每复制一次，末端就要缩短一点。当它的长度减少到一定程度，细胞就会停止分裂。细胞不能够继续复制，从而进入衰老和死亡的程序中，这也正是端粒被称为生命时钟的原因。

这从一个侧面证实了，生命的长度和端粒的长度是相关的：端粒的长度代表了剩余的生命长度，端粒长则生命就长。但是，端粒如何决定生命的长度，这其中到底有哪些具体的机制，穆勒却并不清楚。

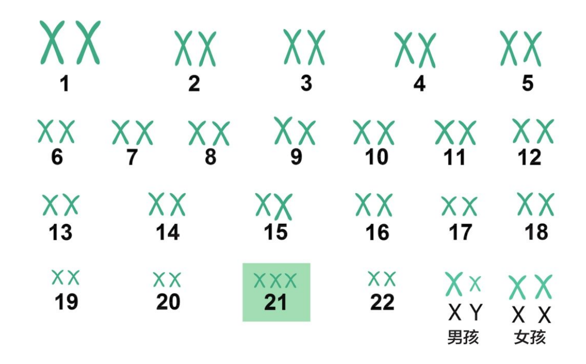
科学家在对早老症患儿的成纤维细胞进行体外培养时，发现他们的端粒磨损的程度要明显高于正常儿童，因此他们的寿命要远远短于

同龄的健康孩子，他们的生命时钟的转速要明显快于正常人。这也从一个侧面证实，端粒磨损的速度关系着我们寿命的长短！



X染色体和Y染色体的端粒

很多遗传性疾病都会影响到人类的寿命。现在，医院会提供很多与遗传有关的疾病检查服务，如孕妈妈们常做的一项产前检查——唐氏筛查。唐氏筛查的主要目的就是分析胎儿患有唐氏综合征风险的高低。唐氏综合征学名为“21号染色体三体综合征”，正常胎儿的第21号染色体是2条，而唐氏综合征患儿体内的第21号染色体是3条。这是一种最早被确定的染色体疾病，一半以上的胎儿会在母体中流产，即使存活下来也会存在寿命短暂、智力低下、发育畸形的情况。




唐氏综合征的核型

## 4.4 拨慢生命时钟

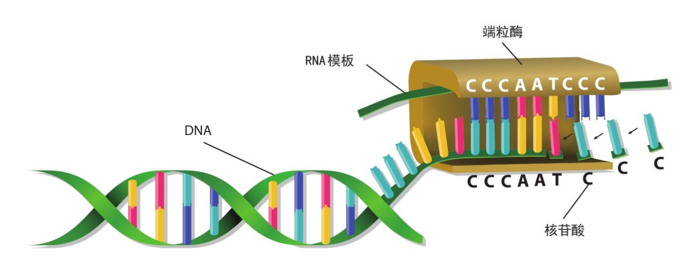
想办法延缓生命时钟的转速，让它能够转动更长的时间，是我们

亟待解决的一件事。一个简单的办法就是将正常衰老的细胞与长生不老的细胞进行比较，了解到底存在什么样的一种机制，让端粒产生变化，是因为缺少了什么样的物质，还是多了什么样的物质，亦或是发生了相应的基因突变？

人类可不可以成为自己生命的主宰，把永生细胞的端粒延长机制引入正常的细胞中呢？

原来，在细胞中有一种酶专门负责端粒的延长，我们将它命名为端粒酶。端粒酶能够以自己的RNA为模板，反转录出端粒序列，以补充磨损的端粒，从而保持端粒的原有长度。在正常的细胞中，端粒酶是没有活性的，或检测不到端粒酶的活动迹象。如果说端粒的长度代表了细胞寿命的长短，那么端粒酶就控制着细胞寿命的长短。

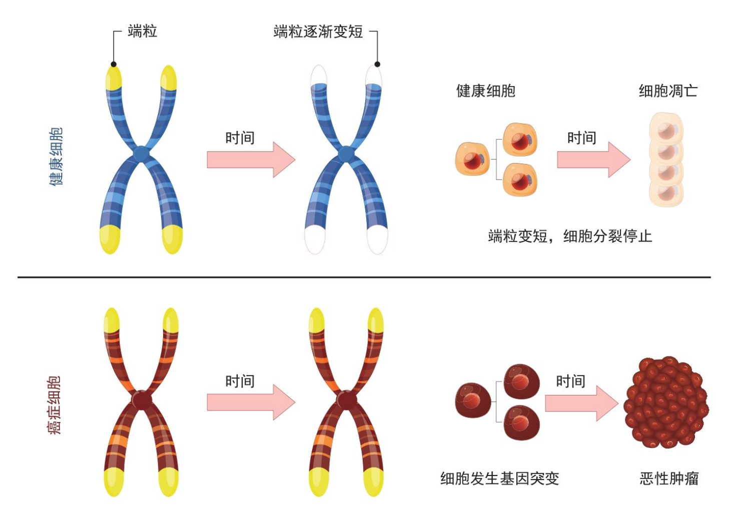
大自然的复杂和精密远远超越人类的想象！很多时候，我们自以为已经掌握了大自然的规律，实际上，却往往只是了解到一点皮毛，真正的规律仍然是“近在眼前，却远在天边”。




端粒酶的作用机制

端粒酶被发现以后，人类似乎找到了一把打开生命时钟外壳的螺丝刀，可以用它来拨慢生命时钟的指针。但是，事情远没有想象的那么简单，结果出乎很多人的预料！

2010年，哈佛大学的肿瘤医生罗纳德·德宾霍（Ronald Debinho）在动物体内进行了大胆尝试，他计划通过激活端粒酶反转录酶，让小鼠“返老还童”。实验起初表现出一定的效果，但是不久很多实验小鼠患上了癌症。




健康细胞与癌症细胞的端粒

斯坦福大学的研究人员也进行过相关的体外实验，他们将编码TERT的mRNA改造后送入人体细胞中，发现人体细胞的端粒会快速有效地延长。如果把TERT基因导入表皮细胞中，端粒可以延长1000个左右的碱基，可以让细胞的复制次数增加40次以上。这大大增加了在体外进行药物测试和疾病建模时细胞的可用性，但是相关的人体实验却迟迟没有展开，一方面是考虑到实验的安全性问题；另一方面是担心出现伦理和道德上的风险。

实际上，染色体末端的端粒具有计数器的作用，在人为改变端粒酶的表达方式时，细胞分裂的计数器被扰乱，细胞分裂就会失控，细胞很容易癌变。只有我们可以操纵端粒酶，使它不诱导细胞向着癌细胞的方向发展，才能对延长人类寿命起到积极的作用。

## 4.5　裸鼹鼠的启示

裸鼹鼠是一种形态极其丑陋的啮齿类动物，看上去就像生化灾难中的变异生物。由于长期生活在地下，裸鼹鼠的眼睛高度退化，几乎丧失了视觉，它的皮肤表面几近无毛，在身体两侧从头到尾长着40余根触须，用来辨别方向和寻找猎物。但它却被科学家高度赞誉：“它

的基因密码可以揭开人类长寿的基因宝盒。”裸鼹鼠的寿命可达30年，大概是普通家鼠寿命的10倍。30岁的寿命也许让很多人觉得不以为意，但是如果换算一下，裸鼹鼠的寿命相当于人类活到500岁。裸鼹鼠为何能够如此长寿呢？

科学家卡尔·罗德里格斯（Carl　Rodriguez）研究发现：裸鼹鼠的细胞因子具有保护体内蛋白质酶的功能。人类在通过酶处理体内存在的垃圾如代谢废物时，自身的蛋白质也会受到相应的损伤，最终导致细胞的死亡，这就相当于日常的生活用品存在磨损一样。裸鼹鼠的细胞因子可以有效地保护垃圾清扫工具——蛋白酶的活性，这样就延缓了衰老的速度。



裸鼹鼠

另外，裸鼹鼠还有一个值得关注的特点，它从来不会罹患癌症。2013年顶级学术期刊——美国《自然》（*Nature*）杂志上发表了一篇关于裸鼹鼠的研究文章，研究发现在裸鼹鼠体内存在着一种叫作透明质酸的物质，这种物质在细胞表面大量聚集，使得细胞之间的联系变得相对敏感，当细胞接触过于紧密时，透明质酸就会发出指令，让细胞停止分裂，从而阻止了癌细胞的发生。

那么人类可以和裸鼹鼠学习吗？可以寻找到抵抗衰老的秘诀吗？

有人说，体积小的物种衰老得慢，也有人说寒冷地区的人衰老得慢，还有人说经常运动的人衰老得慢……我们不要局限于事物的表象，应该拨开迷雾见天日，只需要看一项最主要的指标——生物体代

谢的速度。如果代谢的速度比较快，那么这个物种的寿命便较短；反之，这个物种的寿命便较长。乌龟的长寿，不是因为它的不运动，而是它代谢的速度缓慢。冬眠的动物除了必要的生理活动以外，其他的代谢也基本上处于静止状态。

衰老的速度、寿命的长短不能简单地以运动和静止来衡量，而是应该看代谢的速度。此外，我们还提倡长期、有规律、适度的运动，这也有助于提升人体免疫力，延缓衰老。

有人认为，适当地节食可以延缓衰老，提高人类寿命。科学家做过动物实验，一组小鼠提供充足的食物，另一组小鼠适当地限制饮食。最后，限制饮食的小鼠的整体寿命要比提供充足食物的小鼠长很多。通过解剖，科学家发现，限食的小鼠平均体重要轻1/4，体内的葡萄糖水平降低，代谢速度也相应减缓。实验证实，在适当范围内降低代谢速度，可以起到延缓衰老的目的。同时，节食应当适度。一些女性过度节食导致患上厌食症，不仅不能变得漂亮，反而因过度消瘦危及生命。那么吃多少算是节食呢？我们可以参照这样一个标准，午餐保持八分饱，不要暴饮暴食，保持适度的饥饿感；晚餐七分饱即可，否则会加重肠胃负担。如此便能够起到节食的效果，而不是一味地不吃饭，整日饥肠辘辘。



衰老的过程

# 5　生命的本质——DNA双螺旋结构的发现

作为20世纪最伟大发现的之一，DNA双螺旋结构的确立成为分子生物学诞生的标志，也让人类进入了分子角度研究生命体的时代。从此以后，分子免疫学、分子遗传学、衰老分子生物学等分支学科如雨后春笋般地诞生，从而加速了生命科学的发展步伐。



DNA双螺旋结构

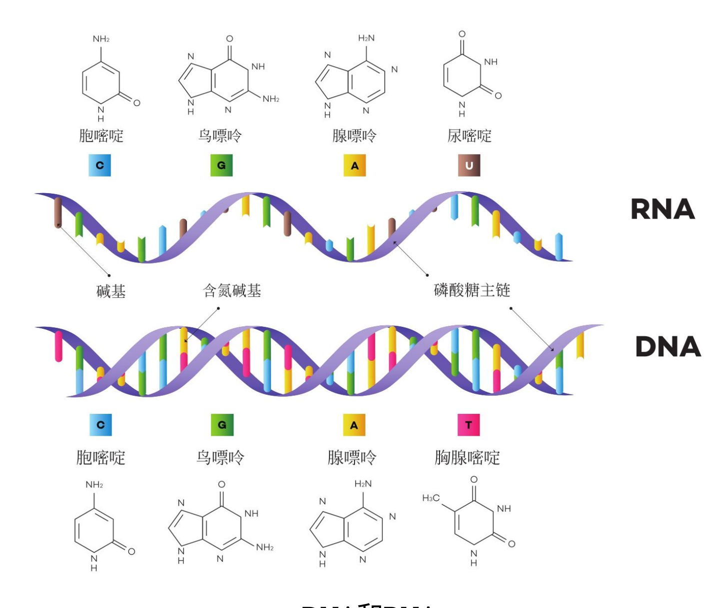
## 5.1　“四核苷酸”假说

组成生物体的主要物质包括糖类、脂类、蛋白质及核酸。在人类并未弄清楚遗传物质的本质之前，这4种分子都是潜在的备选项。刚开始的时候，糖类、蛋白质和核酸都是遗传物质有利的竞争者，但是“四核苷酸”假说的提出，让本应该是正确答案的核酸第一个被淘汰出局。这究竟是怎么回事呢？

说到这里，我们首先得介绍一位重要的科学家，这就是被誉为核酸研究之父的列文（Levene）。

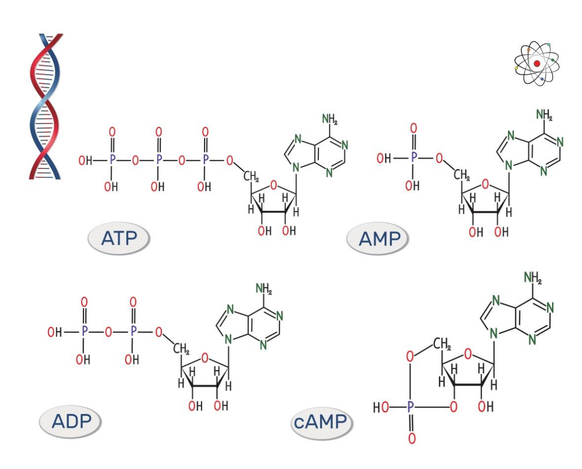
1869年2月25日，列文出生于立陶宛的萨格尔。4岁时，他随父母举家移居到俄国的圣彼得堡，在那里他度过了自己的少年时光。中学毕业后，列文进入帝国军事医学院学习，并在23岁时获得了博士学

位。学习期间，他对生物化学专业产生了浓厚的兴趣。1893年，列文全家迁往美国，在美国期间他从未间断过在生物化学方面的研究。列文后来到欧洲的波恩大学、慕尼黑大学进修，在进修期间，他结识了很多生物化学研究方面的权威人物。列文跟随柯塞尔（Kossel）学习核酸化学、跟随费歇尔（Fischer）学习糖类化学。经过严谨的科研训练，列文的科研水平得以迅速提高。学成归来的列文在1905年被洛克菲勒医学研究所（现在的洛克菲勒大学）聘为助理研究员、化学部主任。列文在这个职位上一直工作到退休。




DNA和RNA

作为美国科学院院士、美国生物化学学会的创始人，列文一共发表了700余篇研究论文，荣获了美国化学会（The American Chemical Society， ACS）的吉布斯奖和纽约地区的尼尔科斯奖章，他在核酸化学领域做出了重要贡献。1868年，“核素”这一概念被提出来之后，柯塞尔的研究组通过大量的反复的实验，证明核酸是由碱基、磷酸和糖类组成的。当时，将取自胸腺的核酸称为胸腺核酸（DNA）；将取自酵母的核酸称为酵母核酸（RNA）。




磷酸腺苷核苷酸

1909年，列文在洛克菲勒医学研究所用酸水解肌苷酸，得到了次黄嘌呤和核糖磷酸；如果改用碱水解肌苷酸，则会得到肌酐和磷酸盐。在此基础上，列文进一步提出了“核苷酸”的概念，并认为核酸是以核苷酸为基本结构单位的。1909年，列文和雅各布斯（Jacobs）通过水解酵母核酸得到了肌苷和鸟苷，然后继续在温和的反应条件下水解，得到了一种结晶的五碳糖，首次证明酵母核酸中的五碳糖是D-核糖。因此，人们将所有的酵母核酸称为核糖核酸。1929年，列文继续用酶解的方法来处理胸腺核酸，得到的居然是脱氧核苷，经过短暂的稀酸处理，他获得了D-2-脱氧核糖的晶体。因为它在酸性环境中极其不稳定，包括柯塞尔在内的多位科学家用酸水解胸腺核酸的方法均无法制得它，所以大家一致认为胸腺核酸的糖是六碳糖，而列文通过自己的研究纠正了这一错误观点。

四核苷酸假说模型

列文还纠正了另外一个错误观点。因为之前的核糖核酸是从酵母、小麦胚芽等植物体中分离出来的，而脱氧核糖核酸是从动物组织如胸腺中分离出来的，所以人们普遍地将核糖核酸和脱氧核糖核酸分别称为植物核酸和动物核酸。列文的研究证实了这种观点是错误的，不管是核糖核酸还是脱氧核糖核酸，在动物和植物体内都有可能存在。列文的主要贡献包括：首次提出并明确“核苷酸”的概念，科学地区分并命名核糖核酸与脱氧核糖核酸，提出了核酸的化学一级结构，证明了DNA分子具有高分子量。这一系列成就，让列文在核酸化学研究领域拥有了崇高的威望。

列文在核酸化学领域做出了重要贡献，他的很多研究成果被奉为经典。然而他的部分错误观点也对后来的核酸研究产生了重要影响。其中，影响最大的就是四核苷酸假说，这一假说目前已被证明是错误的，但是在当时，却被人们奉为经典，即使在奥斯瓦德·西奥多·埃弗里（Oswald Theodore Avery）的肺炎双球菌转化实验证明DNA是遗传物质之后，还是有相当一部分人认为四核苷酸假说是完全正确的。艾弗里迫于舆论的巨大压力，不得不对自己的实验持“谨慎”的态度。从这一点可以看出，四核苷酸假说的影响力是多么的巨大。那么这一影响深远的假说是怎样被提出来的呢？

20世纪初，人们都是用较强的酸来提取核酸，核酸在强酸环境下

很容易分解成短的片段。最初，列文等通过实验测得这些短片段的分子量在1500道尔顿左右，这样的分子量说明核酸是个小分子，并且这个小分子的分子量和四个核苷酸的分子量总和大致相当。又经过仔细的实验，列文发现核酸中四种碱基的含量基本相等。于是，这就顺理成章地形成一种结论，即阻碍核酸研究发展几十年之久的四核苷酸假说：DNA分子是仅含有四个核苷酸的小分子，每种核苷酸的数量大致相同。这一错误观点的最大危害在于，它否定了核酸是大分子物质的客观事实，也排除了核酸成为遗传信息携带者的可能性。因为重复的、过于简单的结构很难在遗传信息的传递中发挥重要作用。

1938年，研究出现了转机，列文和施密特（Schmidt）用超速离心法测出DNA的分子量高达20万～100万道尔顿，而非之前测得的1500道尔顿，这就说明DNA是一种大分子化合物，是具有携带遗传信息潜力的。因为列文对四核苷酸假说深信不疑，所以他仅仅对这一假说进行了些许的修正，再次错过了发现正确理论的机会。列文对四核苷酸假说进行的修改是：DNA分子是由相同的四核苷酸单元聚合而成的高分子化合物。这种简单的聚合物虽然在分子量上达到了大分子化合物的标准，但是因为在结构上过于简单，所以无法成为遗传信息的携带者。

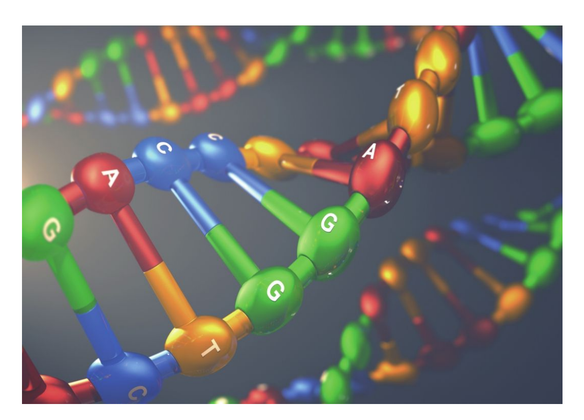
## 5.2　委屈的查伽夫

在生命科学发展的历史上，美国生物化学家查伽夫（Chargaff）做出了巨大的贡献，然而在现实中，很多人却并不了解他的工作。

查伽夫是第一位站出来反对列文的科学家。他质疑列文的四核苷酸假说的正确性，认为这一假说完全排除了核酸作为遗传信息携带者的可能性。查伽夫受过传统的科学教育，是一位语言上的天才，据他自己描述，他可以熟练地使用15国语言。同时，他也是一位有着鲜明个性的科学家，比如查伽夫常说自己是误打误撞地走入了科学研究的殿堂。他宣称，对于生物化学专业，他始终是一个门外汉，是一个旁观者。在看到艾弗里的研究论文之后，查伽夫决定研究DNA。在一开始，检测和精确测量复合物的方法刚刚出现，查伽夫立刻将这种方法

运用在DNA测量上。通过几年时间的持续摸索，1949年他和同事一起发现了一种奇特的现象：四种不同的碱基在DNA中成比例出现，在相同物种的所有组织中，这种比例是恒定的，但是不同物种之间的差距却很大。1950年，查伽夫写了一篇综述，详细地批判了列文的四核苷酸假说，文章中有这么一段话：“然而值得注意的是，这不是偶然的，还没法做出结论。就是说在所有测量过的DNA中总嘌呤和总嘧啶的摩尔（即分子对分子）比值，以及总腺嘌呤对总胸腺嘧啶、总鸟嘌呤对总胞嘧啶的比值都很接近于1。”查伽夫将这个结论告诉了前来拜访他的詹姆斯·杜威·沃森（James Dewey Watson）和弗朗西斯·哈里·康普顿·克里克（Francis Harry Compton Crick），无意之中对DNA双螺旋结构的发现起到了推动作用。1952年5月的最后一个星期，查伽夫与沃森和克里克碰了一次面，此时的查伽夫已经是哥伦比亚大学的正教授，而沃森和克里克还是两个不出名的年轻人。查伽夫的想法给沃森和克里克以极大的启示。9个月后，沃森和克里克构建了DNA分子的双螺旋结构，DNA双螺旋结构模型参考了查伽夫关于碱基1：1比例关系的设想，一条链上的腺嘌呤总是和另一条链上的胸腺嘧啶配对，鸟嘌呤总是和胞嘧啶配对。查伽夫在他的回忆录中用了三页纸来描述这次会面：“我似乎是错过了令人颤抖的认识历史的时刻：一个改变了生物学脉搏节奏的变化……印象是：一个（克里克）36岁，他有些生意人的模样，只是在闲谈中偶尔显示出才气；另一个（沃森）24岁，还没有发育起来，咧着嘴笑，不是腼腆而是狡猾，他没说什么有意义的话。”查伽夫接着写道：“我告诉他们我所知道的一切。如果他们在以前知道配对原则，那么他们就隐瞒了这点。但他们似乎不知道什么，我很惊讶。我提到了我们早期试图把互补关系解释为，假设在核酸链中，腺嘌呤总挨着胸腺嘧啶，胞嘧啶总挨着鸟嘌呤……我相信，DNA双螺旋结构是我们谈话的结果……1953年，沃森和克里克发表了他们关于双螺旋的第一篇文章，他们没有感谢我的帮助，并且只引用了我在1952年发表的一篇短文章，但没有引用我1950年或1951年发表的综述，而实际上他们引用这些综述才更自然。”从文字中能够深深地感受到查伽夫的不满。实际上，他直爽的性格让他在沃森和克里克发表DNA双螺旋结构后没多久，就直接给克里克写了一封信，责备他们没有适当地引用他的工作。查伽夫一个最大的问题在于：他

把DNA考虑成单链，而没有考虑分子是双链的可能性。如果没有双链作为前提，那么即使在知道碱基比例的情况下，也很难构建出这种DNA双螺旋的结构模型。



DNA碱基互补配对模型

不过，从客观上说，查伽夫在DNA双螺旋结构的发现上还是起到了积极的作用。查伽夫最先发现了DNA碱基互补配对规则，但是他为什么没有率先发现DNA双螺旋结构呢？在查伽夫站出来反对列文的四核苷酸假说的时候，他刚刚发现了碱基互补配对规则，也应该意识到核酸是最重要的遗传物质。因此在1950年和1951年，他连续发表了两篇综述，分别介绍了四种碱基在生物体组织中的含量，以及在相同组织中碱基的比例。此时，查伽夫还未看到过DNA的衍射结构图，当然也就不可能从中看出DNA的螺旋结构。在后来研究DNA结构的时候，与威尔金斯（Wilkins）、沃森、克里克等不同，查伽夫没有选择他们都心仪的三螺旋结构，而是固执地认为DNA应该是单螺旋结构。沿此思路，即使结合碱基互补配对规则，查伽夫也无法构建出正确的DNA结构。如果单螺旋要符合碱基互补配对规则，那么这个单螺旋就要发生小肠绒毛状对折，并形成一个个不规则的弯曲。这显然是不合理的，最后他放弃了这一结构模型。

## 5.3　结构学派的圣地

卡文迪什实验室是在科学史上有着重要地位的实验室。DNA双螺

旋结构的发现者沃森、克里克以及与DNA双螺旋结构发现直接或者间接相关的科学家——威尔金斯、佩鲁茨（Perutz）、威廉·劳伦斯·布拉格（Willian Lawrence Bragg）等都与这里有过交集。

1936年，化学家佩鲁茨来到剑桥大学，从事血红蛋白晶体X射线衍射工作的资料收集，在卡文迪什实验室主任、晶体学奠基人、诺贝尔奖获得者布拉格的帮助下开展工作。布拉格进行理论研究，而佩鲁茨进行实验验证，两人合作，一起研究复杂的分子晶体结构。此时，克里克正在佩鲁茨的领导下进行蛋白质晶体结构的研究。

在卡文迪什实验室，大家对于蛋白质的研究兴趣远高于对核酸的研究兴趣，可能就是受列文四核苷酸假说的影响。沃森在《双螺旋：发现DNA结构的故事》一书中写道："我到剑桥以前，克里克只是偶尔想到过DNA和它在遗传中的作用。这并不是因为他认为这个问题没有什么趣味，恰恰相反，他舍弃物理学而对生物学发生兴趣的主要原因是，他在1946年读了著名理论物理学家埃尔温·薛定谔（Erwin Schrödinger）写的《生命是什么》。"当时，人们普遍认为基因是特殊类型的蛋白质分子。然而艾弗里的实验让人们意识到DNA可能是携带遗传信息的载体。此时的克里克并没有打算研究DNA，毕竟他在这个领域已经工作了两年，并且在卡文迪什实验室，大家对DNA的研究兴趣都不大，同时组建一个新的研究小组也需要两三年时间，所以克里克选择继续研究蛋白质结构。与此同时，生物学家威尔金斯的课题组正在进行DNA结构的研究，这个课题组非常小。威尔金斯对DNA结构的研究并没有抱太大的希望，只是按部就班地开展着工作，完全没有料想到多年之后，自己会因此和沃森、克里克一起获得诺贝尔生理学或医学奖。罗莎琳德·富兰克林（Rosalind Franklin）当时是威尔金斯的助手，两人在工作中经常争吵，威尔金斯甚至动了解雇她的念头，却没有找到合适的理由。生物学家莱纳斯·卡尔·鲍林（Linus Carl Pauling）也在研究DNA结构，他向威尔金斯索要结晶DNA的X射线衍射照片副本，被威尔金斯委婉地回绝了。

1951年，沃森进入卡文迪什实验室进行博士后研究，主要从事肌红蛋白的研究。在这里，他认识了比他大12岁的克里克。这一次相遇，孕育了生物学史上的一次伟大发现。沃森和克里克相处得相当融

洽，他们志趣相投，并且两人的研究领域正好互补。克里克在X射线晶体学研究上有着很深的造诣，同时还拥有一定的生物蛋白质学知识。沃森来自著名的学术团队——艾弗里的噬菌体小组，拥有丰富的噬菌体实验工作经验和细菌遗传学研究背景。克里克是一位很有个性的人，可能有些过于自我和狂妄自大，他的性格影响了他与其他人之间的合作，然而沃森却能够包容他的这个缺点，因为他更看重克里克的工作能力和对科学的热情。沃森在《双螺旋：发现DNA结构的故事》一书中提及，克里克虽然从来不知道谦虚，但是自己和他很谈得来，同时他认为克里克是一位在当时就懂得DNA比蛋白质更为重要的人。

## 5.4　最美的DNA双螺旋照片

沃森和克里克能够获得1962年的诺贝尔生理学或医学奖，与另外一位科学家的功劳密不可分，她就是英国著名生物物理学家罗莎琳德·富兰克林。但是现在人们已经很少提及这位女科学家的贡献了，这是不公平的，也是对历史的不尊重。

1920年7月25日，富兰克林出生在英国伦敦的一个犹太家庭中，她的父亲是著名的商业银行家。令人遗憾的是，她在38岁时就早早地离开了人世。如果她没有早逝的话，那么1962年的诺贝尔生理学或医学奖的获奖名单上就会出现富兰克林的名字。在女性科研地位十分低下的当时，富兰克林能取得这样的成就，付出了比其他男性科学家更多的努力。

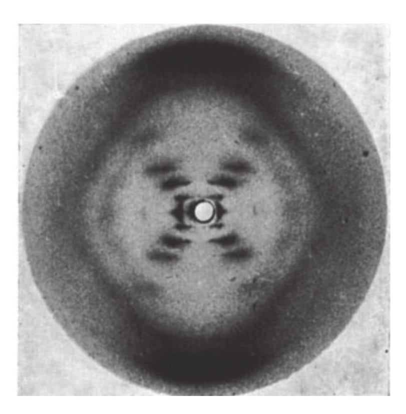


罗莎琳德·富兰克林

少年时代的富兰克林对物理、化学产生了浓厚的兴趣，她18岁进入英国剑桥大学，21岁获得了物理化学专业的自然科学学士学位，25岁获得了剑桥大学博士学位。1947—1950年，她在罗纳德·诺里什（Ronald Norrish）手下从事研究工作。1950年，她受聘于伦敦大学国王学院，从事蛋白质晶体X射线衍射研究。富兰克林任职于伦敦大学国王学院，既是幸运的，也是不幸的。幸运的是，她在这里拍摄出DNA晶体X射线衍射照片。这张照片促进了沃森和克里克构建出DNA双螺旋模型。不幸的是，富兰克林因罹患癌症而离开了人世，这与她长期从事X射线衍射工作有着密切的关系。长时间、大量地接触X射线使她的身体细胞发生突变，从而引发了癌症。对于富兰克林的评价，不同的人有着不同的看法。

沃森在《双螺旋：发现DNA结构的故事》中写道："她学术思想保守、脾气古怪、难以合作、对DNA所知甚少。"1975年，美国作家安妮·塞伊尔（Anne Sayre）出版了《罗莎琳德·富兰克林和DNA》一书，书中展现了一个正直勇敢、宽宏大量，对科学执着、富有激情的女学者形象。无论评价如何，她对发现DNA双螺旋结构的贡献都是无法抹杀的。富兰克林在实验器材和实验样品的处理上下过一番苦功夫。她改进了X射线照相机，使其能够捕捉到像针一样细的光束，并找到了更为合适的方法来排列DNA的绒毛状纤维。1951年11月21日，在伦敦举行的核酸结构学术讨论会上，富兰克林率先展示了一幅DNA结构X射线衍射照片，这是她拍摄的最清晰的一张照片，她使用

的样品是萃取自小牛胸腺的纯DNA样品。



富兰克林拍摄的DNA结构X射线衍射照片

1952年，富兰克林拍摄出了极其清晰的A型和B型两种DNA结构式X射线衍射照片，其中B型的那张照片为日后DNA双螺旋结构的解析提供了实验证据。富兰克林通过不断地改变DNA绒毛纤维周围的空气湿度，使DNA分子在A型和B型之间不断转换。当纤维周围的空气达到75%的相对湿度时，DNA分子就会转变成干燥状态的A形态；当相对湿度上升到95%左右时，DNA分子就会伸长25%，成为B形态。1953年1月，威尔金斯将这张照片展示给了沃森和克里克。沃森在回忆时说道："看到这张照片时，我不禁兴奋地张大了嘴巴，脉搏也剧烈地跳动起来。"

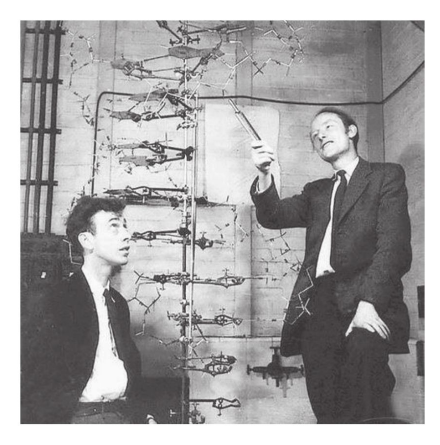
1953年2月24日，富兰克林在研究笔记中记录了DNA分子三螺旋结构的构象，虽然这种三螺旋结构是错误的，但是它已经很接近最终的答案了。3月17日，她完成了关于DNA结构的论文草稿，她推断出DNA每10个碱基为一个周期，距离为34埃，螺旋直径为20埃，这些数据为沃森和克里克提出具体的双螺旋结构模型提供了实验依据。

富兰克林推算出DNA是双链同轴排列的螺旋结构、磷酸根基团和脱氧核糖在螺旋外侧、碱基在螺旋内侧，测定了DNA螺旋体的直径和螺距。1953年年初，DNA分子结构的基本数据已经得到解析，但是尚未形成合理的结构模型。1956年夏天，富兰克林经历了好几次剧烈的疼痛，经检查，她得了卵巢癌。富兰克林在接下来的两年时间里动了

三次手术，还尝试着接受了一些实验性的化学疗法。富兰克林于1958年去世，年仅38岁。1962年的诺贝尔生理学或医学奖颁给了沃森、克里克和威尔金斯，以表彰他们在DNA分子研究方面的贡献，因为他们发现了核酸的分子结构及其对遗传信息传递的重要性。因为诺贝尔奖不颁给已经去世的科学家，所以富兰克林没能获此殊荣。2002年，为了纪念她，英国皇家学会特地设立了富兰克林奖章。

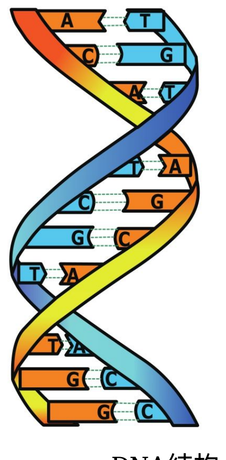
## 5.5　通向生命本质的阶梯

沃森和克里克的合作可以说是生物学史上一个划时代的事件，虽然两人之前的研究领域并没有交集，但是他们之间的合作却碰撞出最耀眼的火花。1953年是生物学史上极有成就的一年，也是分子生物学的诞生之年。从此，人类正式步入分子生物学时代，植物学、动物学、细胞生物学、生物化学等生物学分支学科的科学家们都纷纷展开了分子尺度的研究。沃森和克里克在卡文迪什实验室相识，美国化学家鲍林的儿子彼得和另外一位科学家多诺休（Donohue）也在卡文迪什实验室，两人和沃森、克里克逐渐有了工作上的往来。1952年12月的一次交谈中，彼得告诉沃森，鲍林在给他的一封家信中提到自己已经完成了DNA分子结构的构建。这让沃森感到非常紧张，他和克里克已倾注了多年的心血，如果被抢先发现，那么他们所有的努力都将化为乌有。沃森于是催促彼得给鲍林写信，希望能够获得论文的复印件。不久，论文《一个核酸结构的建议》寄到了剑桥。这个模型结构是错误的，和之前被他们否定的错误模型有几分类似，沃森和克里克都松了一口气。有机分子有多种同分异构体，因此沃森和克里克带着草图去请教化学家多诺休。多诺休指出，草图中的碱基构型的烯醇式应该改为酮式异构体。沃森回忆道："1953年2月20日星期五的这一刻，我们彻底明白了碱基在分子内部靠氢键的专一性来配对。"1953年4月25日，《自然》杂志发表了沃森和克里克发现DNA双螺旋结构的论文。这篇论文并不长，只有薄薄的一页，但是这薄薄的一页纸却改写了生物学的历史，开创了现代分子生物学的研究先河。这篇解读人体遗传物质的论文被称为"人类有史以来伟大的50篇论文之一"。



沃森、克里克与双螺旋结构模型

客观地说，沃森和克里克的科研之路也是充满坎坷的。沃森和克里克受到富兰克林那张衍射照片启发，开始着手构建DNA的结构模型。他们首先构建的是DNA三螺旋结构，也就是三条不同的DNA链相互缠绕在一起形成的螺旋模型。富兰克林犀利地指出，他们的模型在结构上有很多缺点，比如结构不稳定、含水量与实际测量结果间存在很大的误差等。因此，这一模型刚面世就宣告失败。这次模型构建的失败，对两人来说都是一次极大的打击，让他俩都有些心灰意冷。在随后半年里，克里克回归到自己的蛋白质课题研究中，沃森也开始继续研究烟草花叶病毒。关于DNA结构模型构建的事情就被暂时搁置了起来。1952年6月的一天，克里克在一次茶会上遇到了年轻的数学家约翰·格里菲斯（John Griffith）。格里菲斯告诉克里克，他已经完成了DNA中碱基互补吸引配对的计算。这次深入的交谈又一次激起了克里克继续研究DNA结构的热情，克里克立刻联系了自己的老搭档沃森。也许是源于对DNA结构的痴迷，抑或是对于未知结构探索的渴望，沃森爽快地答应了克里克的邀请。两人再一次联手，开始迎接新一轮的挑战。1952年7月，克里克和沃森拜访了生物化学家查伽夫，查伽夫明确地告诉沃森和克里克，不同种类的碱基在总量上完全符合1：1的比例关系，也就是说四种碱基分别是互补配对的。这意味着两人距发现DNA双螺旋结构仅剩最后一层窗户纸了。




DNA结构



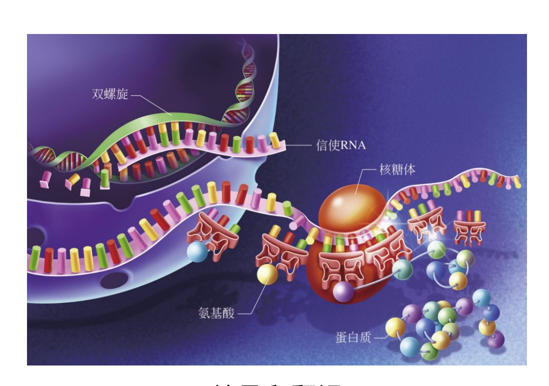
1953年1月，沃森再次来到伦敦大学国王学院，拜访了威尔金斯。从威尔金斯的口中，他听到了富兰克林报告的全部内容。而此时的威尔金斯依然偏爱DNA三螺旋结构，始终认为这个模型最符合DNA的密度值。威尔金斯的研究因此走入了死胡同。在查伽夫规则的影响下，沃森和克里克彻底摒弃了DNA三螺旋结构的思路，开始思考双螺旋结构是否更适合。按照双螺旋结构建立模型的过程出乎意料地顺利，双螺旋结构完美地解释了包括稳定性、含水率、衍射图在内的绝大多数问题。因此，这一结构逐渐地获得了沃森和克里克的认可。1953年春夏之交，沃森和克里克一共写了四篇关于DNA结构与功能方面的论文。第一篇顺利地发表在《自然》杂志上。紧随其后，威尔金斯、艾力克·斯托克斯（Alec Stokes）、贺伯特·威尔森（Herbert Wilson）、富兰克林和戈斯林也发表了两篇论文。五个星期后，沃森和克里克又在《自然》杂志上发表了第二篇论文，这次的主题是讨论DNA双螺旋的遗传学意义。这两篇文章奠定了他们在分子生物学研究中的鼻祖地位。有人说沃森和克里克发现DNA双螺旋结构就像是哥伦布发现了新大陆。其实两者之间存在着极大的不同，生物学研究除了自身的实力之外，还受到其他很多因素的制约，包括实验经费、实验

技术、运气等。

# 6　从直线到三角——生命公式的完善

DNA双螺旋结构建立之后，人类进入了分子生物学时代。随后，科学家们就开始思考遗传物质是如何将这些蕴含在双螺旋结构中的信息表达出来的呢？它遵循着什么样的法则呢？

## 6.1　遗传中心法则的发现



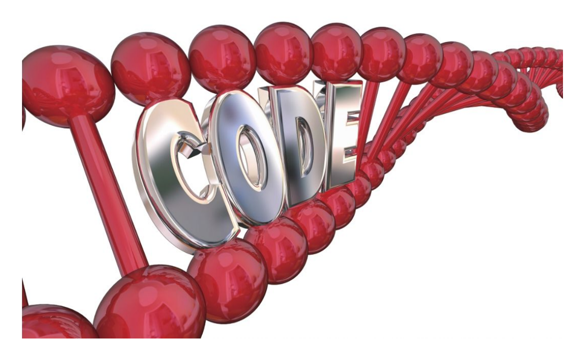

转录和翻译

DNA双螺旋结构确立后，生物学家、信息学家都在想生物体内是不是像传递情报一样，存在一整套密码系统，通过密码的读取，把遗传的信息解读出来呢？

既然遗传信息是存储在DNA上的，生物的性状又是通过蛋白质表现出来的，那么在蛋白质和DNA之间一定存在两种联系：第一，在DNA的核苷酸链上，碱基的排列顺序就决定了基因的遗传信息；第二，基因携带的信息除了代表一种给定的多肽的一级结构外，不包含其他的信息。那么，在DNA和蛋白质之间究竟有没有其他联系的物质，起联系作用的机制又是什么？当时却并不清楚。

1954年，生化学家乔治·伽莫夫（George Gamow）率先提出了一种遗传密码方案，他认为，在DNA的多核苷酸链上存在着一组组以相邻的三个核苷酸碱基作为一种氨基酸编码的密码。这种三联体密码是

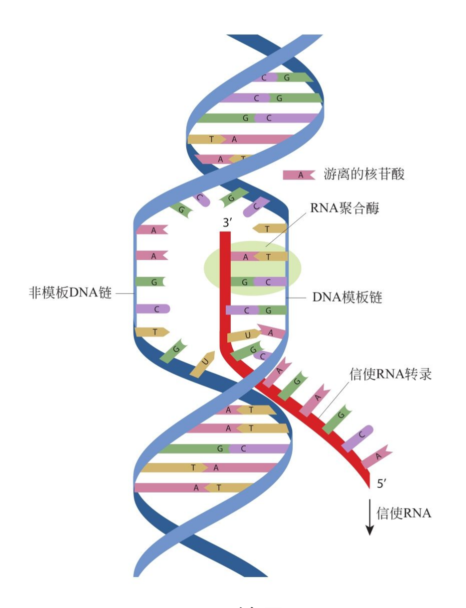
有重叠的，因此一个氨基酸可能存在几种不同的密码。这是第一种以三联体形式作为遗传密码的解读方案，具有一定的先进性。但是，伽莫夫的这一方案仅仅是理论上的猜测，并没有进行实验验证。



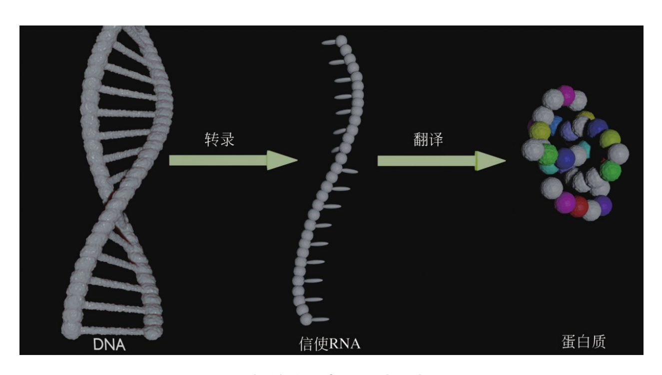

遗传密码

20世纪50年代，在研究DNA的同时，还有一部分科学家致力于研究RNA和蛋白质之间的编码关系，布拉舍特（Brachet）和卡斯帕森（Caspersson）就提出了RNA控制蛋白质合成的观点。但是随后的发现证实，RNA合成蛋白质是在核糖体上进行的，即使DNA在受到酶的破坏之后，依然会有蛋白质的合成。因此，蛋白质的合成是受细胞质中的RNA直接控制的，可能遵循着从DNA到RNA再到蛋白质的过程。

1955年，克里克发现三联体密码的长度约为10埃，而氨基酸分子的长度为2～3埃，存在着明显的差异，因此两者之间一定还存在着一些中介物。在不清楚中介物是什么的情况下，克里克提出了适配器学说。这一学说认为：氨基酸并不和模板直接结合，而是首先和一种特异的受体分子结合。这样，模板和氨基酸分子在体积上就能够完全匹配了。




DNA转录




遗传的中心法则

1957年，美国生物化学家蒂姆·霍格兰（Tim Hoogland）在大鼠的提取液中发现一种RNA能够和氨基酸相结合，这一发现证实了克里克的配适器学说是完全正确的。

1958年，克里克根据实验结果提出了中心法则，认为遗传信息只能从核酸流向蛋白质，传递是单方向进行的。虽然这一观点后来被证实并不完全正确，但是在当时却有着重要的意义。 1961年，弗朗索瓦·雅各布（François Jacob）和莫诺（Monod）把这种能够将遗传信息从DNA转移到核糖体上的物质称作“信使”。他们提出每个DNA基因

的核苷酸顺序都是转录在RNA分子上的，由此确定了RNA的信使作用。

## 6.2 三联体密码子的确定

1961年是遗传三联体密码研究取得重要进展的一年。克里克和悉尼·布伦纳（Sydney Brenner）进行了一项重要实验，解决了遗传密码传递信息的问题。他们利用T4噬菌体的γII基因做材料，经原黄素类化学诱变剂处理后，应用移码突变的方法进行验证。实验是这样进行的：在一条多核苷酸链的相邻的两个核苷酸中间，插入一个由核苷酸引起的突变，会使译码过程中读码的起点移位，结果在肽链之间插入一段不正确的氨基酸。

如果在该噬菌体的DNA中再减去一个碱基，或者再加上两个碱基，那么就会让编码蛋白质的结果恢复到原来的样子，不再有突变发生。这说明核酸的密码是以三个核苷酸为一组所组成的。

克里克和布伦纳根据实验得到了三条正确的结论：①信息从基因的一端不重复地连续读出，信息阅读的对错取决于信息的读取起点；②信息的读出以三个核苷酸为一组；③大多数的三联体密码可以决定一个氨基酸的合成，只有少数是没有意义的，因此很多氨基酸都有一个以上的同义码。1961年夏天，美国生物化学家尼伦伯格（Nirenberg）和德国生物学家马太（Matthaei）取得了突破性进展。他们建立了一个无细胞系统，把编码氨基酸的mRNA引入无细胞系统中，用来指导某一种多肽的合成。当他们把全部碱基都是尿嘧啶（U）的多聚尿苷引入后，产生的都是苯丙氨酸，这说明苯丙氨酸的密码是UUU。随后，生物学家塞韦罗·奥乔亚（Severo Ochoa de Albornoz）和同事进行了一系列破解实验，在一年内弄清楚了多种氨基酸的密码子。1964年，生物学家哈尔·葛宾·科拉纳（Har Gobind Khorana）通过一系列双密码子的交替共聚物实验，确定了密码排列的顺序问题。1966年，克里克根据已经取得的成果，排列出遗传密码表。20世纪70年代，比利时肯特大学的菲尔斯（Fiers）等用MS2噬菌体做材料，对三联体密码进行了验证。他们分析了MS2噬菌体外壳蛋

白中129个氨基酸的顺序，以及与外壳蛋白对应的390个核苷酸的顺序，其结果完全符合遗传密码表上的对应关系。至此，三联体密码系统正式为人们所认同。到目前为止，三联体密码在整个生物界都是适用的，这也从分子水平上证明了有机体遗传信息传递的重要性。

遗传密码表

| 第1密码子 | 第2密码子 | | | | 第3密码子 |
|---|---|---|---|---|---|
| | U | C | A | G | |
| U | 苯丙氨酸（Phe） | 丝氨酸（Ser） | 酪氨酸（Tyr） | 半胱氨酸（Cys） | U |
| | 苯丙氨酸（Phe） | 丝氨酸（Ser） | 酪氨酸（Tyr） | 半胱氨酸（Cys） | C |
| | 亮氨酸（Leu） | 丝氨酸（Ser） | 终止密码子 | 终止密码子 | A |
| | 亮氨酸（Leu） | 丝氨酸（Ser） | 终止密码子 | 色氨酸（Trp） | G |
| C | 亮氨酸（Leu） | 脯氨酸（Pro） | 组氨酸（ His） | 精氨酸（Arg） | U |
| | 亮氨酸（Leu） | 脯氨酸（Pro） | 组氨酸（His） | 精氨酸（Arg） | C |
| | 亮氨酸（Leu） | 脯氨酸（Pro） | 谷氨酰胺（GIn） | 精氨酸（Arg） | A |
| | 亮氨酸（Leu） | 脯氨酸（Pro） | 谷氨酰胺（GIn） | 精氨酸（Arg） | G |
| A | 异亮氨酸（Ile） | 苏氨酸（Thr） | 天冬酰胺（Asn） | 丝氨酸（Ser） | U |
| | 异亮氨酸（Ile） | 苏氨酸（Thr） | 天冬酰胺（Asn） | 丝氨酸（Ser） | C |
| | 异亮氨酸（Ile） | 苏氨酸（Thr） | 赖氨酸（Lys） | 精氨酸（Arg） | A |
| | 甲硫氨酸（met） | 苏氨酸（Thr） | 赖氨酸（Lys） | 精氨酸（Arg） | G |
| G | 缬氨酸（Val） | 丙氨酸（Ala） | 天冬氨酸（Asp） | 甘氨酸（Gly） | U |
| | 缬氨酸（Val） | 丙氨酸（Ala） | 天冬氨酸（Asp） | 甘氨酸（Gly） | C |
| | 缬氨酸（Val） | 丙氨酸（Ala） | 谷氨酸（Glu） | 甘氨酸（Gly） | A |
| | 缬氨酸（Val） | 丙氨酸（ Ala） | 谷氨酸（Glu） | 甘氨酸（Gly） | G |

三联体密码

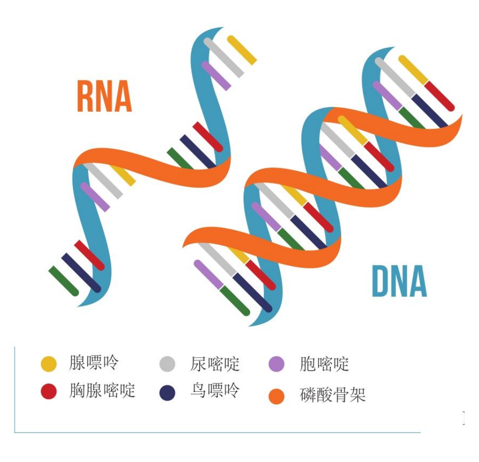
## 6.3　RNA酶的发现

除了中心法则外，“酶”的概念也在不断丰富。长期以来，人们一直认为酶的本质就是蛋白质，RNA酶的发现宣告了酶的本质不是单一的。RNA病毒、RNA反转录酶以及朊病毒的陆续发现表明，在一定条件下RNA和蛋白质都可以作为携带遗传信息的载体。

RNA酶的发现再次证实了生命现象的多样性和复杂性。

中国人对酶的利用已经持续了几千年。早在公元前21世纪的夏禹时期，人们就已学会了酿酒；在3000年前的周朝，人们已开始制作饴糖和酱；在2000多年前的春秋战国时期，人们已经知道用麦曲来治疗消化不良等肠胃疾病。虽然那时候的人并不了解酶的本质，但是酶的利用已经相当广泛。

现代生物化学研究表明，生物的新陈代谢等基本的生命活动都是在酶的催化下通过生物大分子的合成和分解来完成的。生物体中的酶是一种生物催化剂，它通过降低生物的活化能来加速和调节生物体内的生物化学反应。直到20世纪，人类才揭示了酶的本质。1926年，美国化学家詹姆斯·巴彻勒·萨姆纳（James Batcheller Sumner）从刀豆中提取了脲酶并将其结晶，这证明了它具有蛋白质的特性。1930—1936年，诺思罗普（Northrop）和库尼茨（Kunitz）先后得到了胃蛋白酶、胰蛋白酶和胰凝乳蛋白酶的结晶，并证实它们均属于蛋白质。从此，“酶的本质是蛋白质”成为学术界的共识。为此，萨姆纳和诺思罗普在1949年一起获得了诺贝尔化学奖。

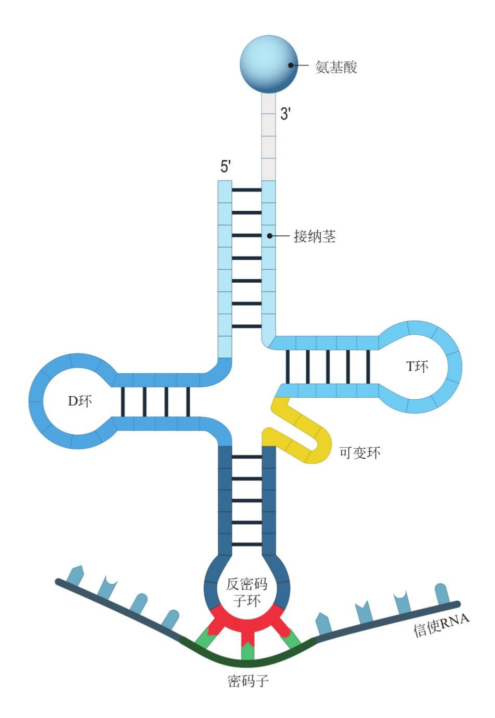



DNA和RNA的区别

虽然绝大多数的生物是以DNA为遗传物质，但一些病毒和噬菌体却是以RNA为遗传信息的载体，这显示了RNA功能的多样性和生命现象的多样性，然而人们似乎很难将RNA的功能与酶联系在一起。

在RNA酶的研究过程中出现了两位重要的科学家。其中一位是美国生物化学家托马斯·罗伯特·切赫(Thomas Robert Cech)，他从小就对自然科学十分感兴趣。1966年，切赫进入格林内尔学院学习，实验室中的一系列设计、观察和解释训练让他喜欢上了生物化学。1975年切赫获得博士学位，随后前往麻省理工学院进行博士后研究，在帕杜的实验室中，他学到了很多生物学知识，对生物学逐渐产生了浓厚的兴趣。1978年，切赫在科罗拉多大学担任教职。在科罗拉多大学，切赫选择了四膜虫作为研究RNA拼接机制的对象。内含子是核酸链上不编码任何蛋白遗传信息的碱基片段。1986年，切赫在四膜虫rRNA前体中观察到一个有395个碱基的线状RNA分子组成的内含子，他将其命名为L19RNA。经过深入研究，切赫发现L19RNA具有类似酶的作用，在一定条件下能够像酶一样以高度专一的方式去催化寡聚核糖核苷酸底物的切割或连接，既有核糖核酸酶活性，又有RNA聚合酶活性。






转移RNA

另一位重要的科学家是西德尼·奥尔特曼（Sidney Altman）。1939年5月7日他出生于加拿大蒙特利尔一个贫穷的移民家庭。最初激发奥尔特曼对科学产生兴趣的是原子弹和元素周期表。大学毕业后，奥尔特曼作为物理学系的研究生来到哥伦比亚大学。从他的专业背景看，似乎学习物理学更能够发挥他的研究特长，但是物理学家伽莫夫看出了他在生物学上的天分，于是把奥尔特曼推荐给科罗拉多大学从事将染料分子插入DNA研究的伦纳德·勒曼(Lenard Lerman)，从而使他与核酸分子打起了交道。在完成吖啶分子对T4噬菌体DNA复制影响的研究后，他加入了哈佛大学的梅塞尔森实验室，研究核酸内切酶在T4DNA复制和重组中的作用。

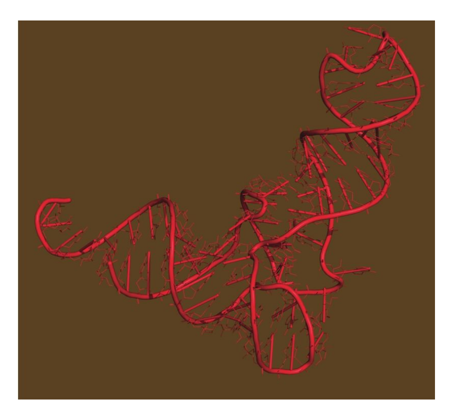
1967—1971年，奥尔特曼在剑桥大学继续从事用大肠杆菌进行tRNA合成的研究。他提取到了纯化的tRNA的前体——一种生物合成tRNA的中间产物。按照生物化学代谢的规律，如果有一个中间产物，那么就意味着存在一种催化生成这种中间产物的酶。据此，他顺利地找到了核糖核酸酶P，它的功能就是切开RNA链上的磷酸二酯

键，释放出最终的tRNA。这种RNA酶在反应中不会被消耗，同时能够加速反应，完全符合酶的特性。切赫和奥尔特曼在不同的实验室用不同的实验材料证明了某些RNA分子具有生物催化的功能，按照酶的定义，可以称之为RNA酶。

RNA酶的发现意味着酶的本质不一定是蛋白质，从而向酶本质的传统观念提出了挑战。不久，越来越多的具有催化自我剪接功能的RNA被发现。到1989年，核酸酶终于得到大家的认可并被写进教材。它的发现者切赫和奥尔特曼也因此荣获了1989年的诺贝尔化学奖。

随着分子生物学研究的不断深入，RNA在生命活动中的重要作用将会越来越多地被揭示出来。生命起源一直是科学家们关注的课题，生物体内的DNA、RNA和蛋白质等生物大分子，在生物的遗传和生命现象的表达中各司其职、相互配合、缺一不可。然而在生命起源的初期，哪一种生物大分子是最早出现的呢？RNA酶的发现说明，RNA在遗传方面的功能更为全面，从携带遗传信息、调节基因表达到催化自我复制，RNA在某些场合中可以不需要DNA和蛋白质而完成自我复制，因此RNA是一种可以独立进行生命表达和遗传的大分子。

《科学》（*Science*）杂志在2000年12月介绍当年的重大科学成就时，把人类基因组工作草图的绘制工作放在了第一位，认为生命可能是源于RNA而非DNA。沃特·吉尔伯特（Walter Gilbert）率先提出了“RNA世界”假说，他认为在生命起源初期，RNA已经表现出了DNA和蛋白质的某些功能特性，同时RNA在生物体的遗传信息等方面还起着承上启下的纽带作用。一些常见的低等生物，如艾滋病病毒、丙型肝炎病毒、埃博拉病毒和烟草花叶病毒等均以RNA为遗传信息载体，因此地球上的生命起源很可能是从RNA开始的，中心法则中DNA、RNA和蛋白质等生物大分子的基本分工应该是生物长期演化的结果。

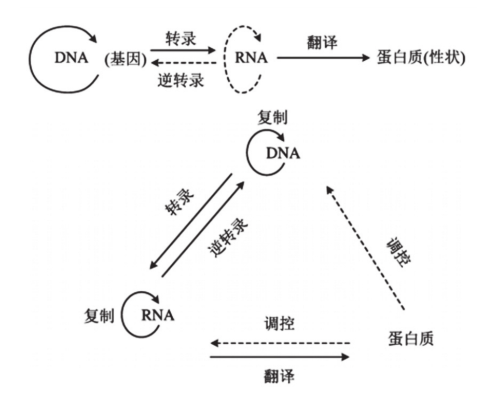


RNA世界

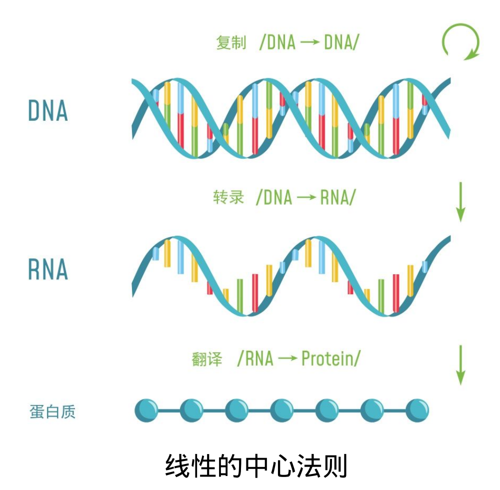
## 6.4　稳定的三角

对于自然界来说，所有具有遗传物质的动物和植物，都遵守着一条内在的定律——被称为生命公式的中心法则。中心法则是克里克于1958年提出的，最初并没有得到学术界的足够重视。1970年，克里克在《自然》杂志上重申：遗传信息既可以从DNA传递到RNA，再从RNA传递到蛋白质，完成遗传信息的转录和翻译过程，也可以从DNA传递到DNA，完成遗传信息的复制过程，但是不能由蛋白质转移到蛋白质或者核酸中。

朊病毒的发现完善了中心法则。从此，生命公式从原先的直线形式变成了现在的三角形的相互关联的形式。



中心法则的完善




线性的中心法则

现在的中心法则呈现闭合的三角关系，顶点之间皆存在着密切的联系。在自然界中，三角形结构是较为稳定的，有很强的抗压能力。然而也不能妄言这就是定论，毕竟科学的发展是没有止境的，人类只有在发展中不断地完善已有的学说，才是对待科学应有的严谨态度！



# 7　盘旋的公路与基因测序

诺贝尔奖获得者杜伯克（Dubbecco）曾说过一句名言：人类的DNA序列是人类的真谛，这个世界上发生的一切都与之息息相关。基因的本质是特定的DNA片段，因此测定DNA序列的重要性不言而喻。要研究DNA中蕴含的巨大遗传信息，就必须了解它的核苷酸序列结构，于是人类开始琢磨如何去测定这样复杂的序列。



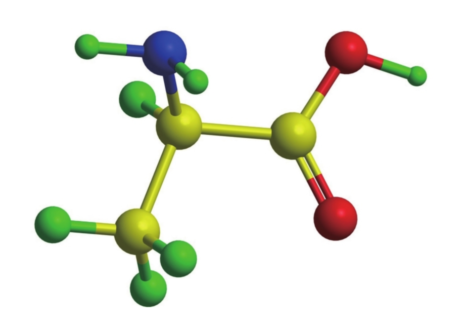
基因的解读

## 7.1　测序蛋白质和RNA

早在沃森和克里克的DNA双螺旋结构建立之前，学术界尚未认定DNA就是遗传物质的载体，反而认为通过测序蛋白质和RNA序列可能更快地解密人类的遗传奥秘。当时，科学界对于遗传物质究竟是什么的争论尚未有明确的答案，那么摸清蛋白质和RNA的分子中蕴含的遗传信息是研究的当务之急，所以在DNA测序技术发明之前,蛋白质和RNA的测序技术就已经出现。

最早被测序的核酸分子是丙氨酸tRNA，这项工作由诺贝尔奖得主美国生物学家罗伯特·W.霍利（Robert W. Holley）完成。1922年，霍利出生在美国伊利诺伊州，他的父母都是教育界人士。1942年，霍利从伊利诺伊大学毕业，1947年他在康内尔大学获得有机化学博士学

位。他一直对生命现象感兴趣，所以最初选择从事自然产品的有机化学研究。随后他的研究方向逐渐由化学向生物学方面转变，开始研究氨基酸和蛋白质的生物合成，最后他的兴趣集中到了RNA分子上，尤其是酵母丙氨酸转移RNA。在这个看似不起眼的小分子上，霍利倾注了10年的心血。他先是把丙氨酸转移RNA提纯出来，然后研究它的结构，最后对这个分子进行测序。霍利使用的测序思路和弗雷德里克·桑格（Frederick Sanger）测胰岛素的思路一样：先对RNA分子进行部分酶切，再用离子交换柱分离RNA片段，对小的RNA片段进行碱基测序。1965年他在《科学》杂志发表了关于酵母丙氨酰-tRNA序列的76个核苷酸的序列测定的文章，就是利用这种技术方法完成的。这一整套的工作为他赢得了1986年的诺贝尔生理学或医学奖。

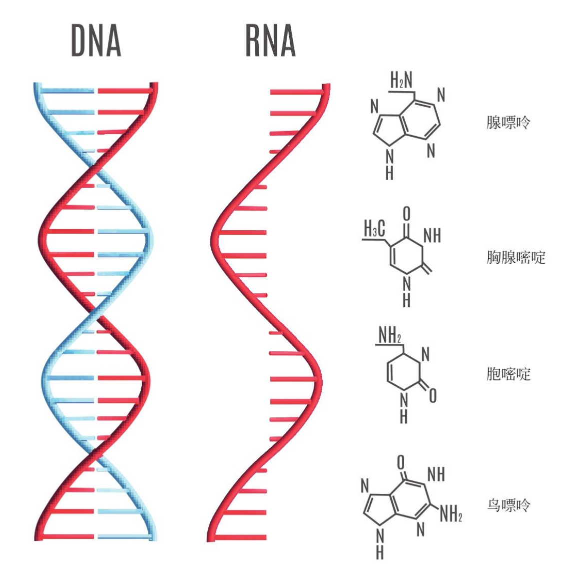


丙氨酸分子结构

20世纪60年代RNA序列测定技术率先发展起来，DNA测序的最早尝试也是借鉴于此方法，称为小片段重叠法，这种方法也正是霍利研究成果的衍生。

另一位对测序方法产生重大影响的人物是英国科学家桑格，下一节，我们将详细介绍这位科学家。桑格研究小组也对RNA测序进行了研究，通过桑格自己偏爱的两项技术——同位素标记和纸层析，他们找到了更快捷的方法。RNA分子特别适合用P32标记，而且双向纸层析比离子交换层析要省力得多。用这种新方法，桑格实验室的G.G.布朗利（G. G. Brownlee）测出120个碱基的5S核糖体RNA，它是当时被测序的最大的RNA分子。对RNA进行测序积累起来的相关实验经验，对桑格研究小组随后研究DNA测序技术起到了重要的铺垫作用。

由于蛋白质和RNA结构较DNA结构简单，进行测序的步骤和难度也相对较小，所以蛋白质与RNA测序技术成为了DNA测序研究的前奏，为DNA测序进行先行的探索，并提出了合理的思路，避免了不必要的重复，为DNA测序研究奠定了基础。




DNA和RNA的区别

## 7.2　“加减法”与“双脱氧法”

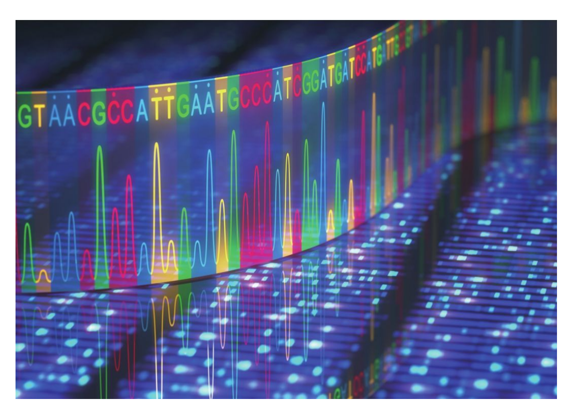
早在20世纪50年代，关于DNA测序技术的研究就已经逐步开始了。包括化学降解法、多聚核糖核苷酸链降解法等，但是由于测定方法不成熟、操作复杂等原因，这些方法并未得到广泛运用。

M.克莱因（Morris　Kilne）曾经说过这样的话：一个人应该有足够敏锐的思想从纷乱的猜测中清理出前人有价值的想法，并且有足够的想象力把这些碎片重新组织起来，从而能够大胆地制订一个宏伟的计划。在DNA测序技术的发展中桑格就扮演了这样一个角色，他完成了最艰难的总结步骤，第一个将DNA测序方法系统化和标准化。

1918年8月31日桑格出生在英国格洛斯特郡的一个富足家庭。他在剑桥大学学习时接触到了生物化学，立刻对它产生了浓厚的兴趣，

从此他将自己的全部精力都投入其中。1943年桑格获得剑桥大学的博士学位，到1951年他一直在学校继续着生物化学的研究工作。1938年桑格开始研究胰岛素，他发明了一种方法用来标记N端氨基酸，此后，他得到了医学研究理事会的赞助并继续进行研究工作。经过10年的不懈努力，桑格在1955年测定了牛胰岛素的蛋白结构序列，为今后在实验室合成胰岛素奠定了基础，因此桑格获得了1958年的诺贝尔化学奖。

从1975年开始，桑格逐渐将自己的精力转移到DNA测序的研究上来。他和科学家库尔森（Coulson）一起发明了“加减法”测定DNA序列。两年之后，他在研究的基础上，继续改进实验的方法，通过引入双脱氧核苷三磷酸形成双脱氧链终止法，使得DNA测序的稳定性得到大幅提升。这种方法通过核酸模板在DNA聚合酶、引物、单脱氧核苷三磷酸存在的条件下进行复制，依靠将双脱氧核苷掺入链的末端使之终止或者引入单脱氧核苷使之继续延长。实验结束后可以得到一系列长短不一的片段，长度相邻的片段分别相差一个碱基，通过比较不同的核酸片段，利用放射自显影的技术就可以一次读出不同的碱基排列顺序。这种方法进行测序实验步骤简单，同时误差较小，所以以后很多与之相关的测序技术的改进都是在此基础上进行的，桑格开辟了最行之有效的测序方法。




DNA测序

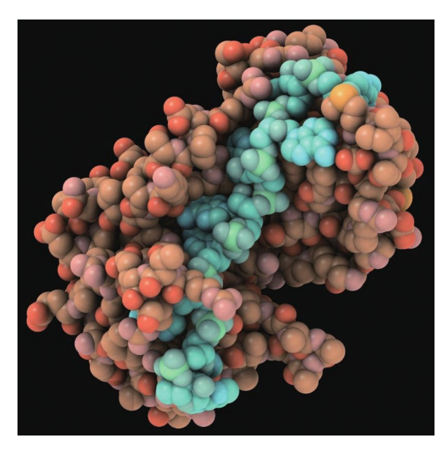
桑格在加减法测序的基础上继续对DNA测序加以改进，1980年他

在此方法基础上设计出双脱氧法，这种测定DNA序列的方法直到现在还被广泛使用。桑格与沃尔特·吉尔伯特（Walter　Gilbert）、保罗·伯格（Paul　Berg）共同获得1980年的诺贝尔化学奖，他成为第四位两次获此殊荣的科学家，这是科学界对其测序研究成果的肯定，也为日后人类基因组序列的揭秘提供了最快捷的方法。

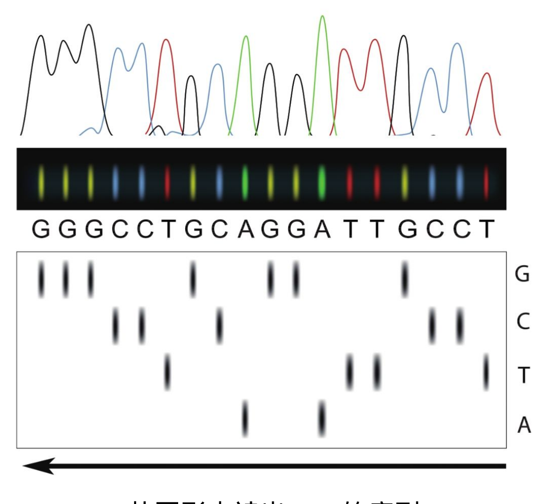
吉尔伯特用不同的思路对DNA测序，同样取得了成功。吉尔伯特1932年3月21日出生在美国波士顿，他的父亲是哈佛大学的经济学家和政府的经济顾问，他的母亲是一位儿童心理学家，总喜欢拿吉尔伯特和他妹妹做试验。在父母的影响下，他和妹妹从小就喜欢阅读。在高中时，他对无机化学产生了兴趣，之后又迷上了核物理学，他经常逃课去国会图书馆翻看相关的书籍。在哈佛大学，吉尔伯特的专业是化学和物理学。在读完一年的研究生后，他转学到了英国剑桥大学，在那里拿到了物理学博士学位。之后他返回哈佛大学发展，两年后成了物理系的助教。在20世纪50年代后期，他开始从事理论物理的教学，不过其兴趣渐渐转到了实验领域。

1960年，沃森等在寻找信使RNA这一把DNA信息传递出去的载体，吉尔伯特也于当年夏天加入这项研究。一条信使RNA的寿命很短，在指导几次蛋白质合成之后就会被降解，降解后的产物被细胞重新利用。这个实验给吉尔伯特的感觉是“抓住了生命活动的一瞬”。

20世纪70年代中期吉尔伯特开始研究快速测定DNA顺序的方法，他们研究出的化学法不需要任何酶，对单链和双链DNA都适用，不受DNA二级结构的干扰，因而被广泛应用于各地实验室。化学法的原理是利用四种不同碱基的特征化学反应除去特定种类的碱基，使链在该处断裂，得到各种长度的同位素标记的片段，在聚丙烯酰胺变性胶上电泳分离，根据长度读出碱基顺序。



蓝绿色的信使RNA与结合蛋白




从图形中读出DNA的序列

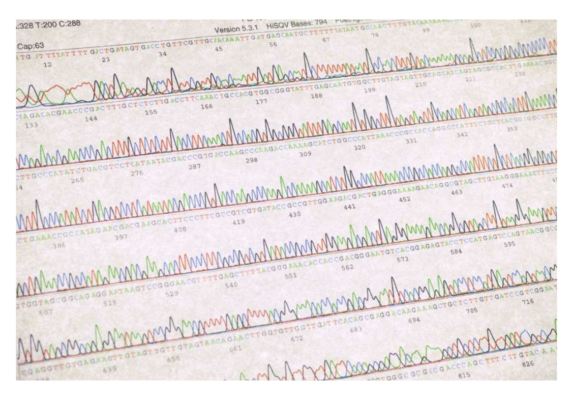
吉尔伯特的化学降解法和桑格的双脱氧链终止法成为当时应用广泛的DNA测序方法，虽然这两种方法在技术和原理上存在很大的差异，但是都生成相互独立的寡核苷酸，这些核苷酸上都带有放射性的标记，通过形成共同的起点，终止于不同的残基，形成了多条不同种长度的寡核苷酸片段。随后对其进行高分辨率的凝胶电泳，将不同的片段分离开来，利用放射自显影技术从胶片上就可以直接读出所需测定的DNA的核苷酸序列。

## 7.3　华人科学家吴瑞

美籍华裔科学家吴瑞教授在测序技术的发展史上也占有一席之地，他提出新的引物延伸的测序序列：先将引物加以定位，然后用此引物来延伸和标记新的DNA。后来桑格的DNA快速测序法和凯利·穆利斯（Kary Banks Mullis）的PCR技术都是以这种测序技术的想法为基础发展起来的。

吴瑞的祖籍是福建省福州市，1928年8月14日生于北京。在家庭环境的熏陶下，他接触到以此为终生职业的生物化学。1948年7月吴瑞跟随母亲来到美国与父亲团聚。抵美后的暑假期间，吴瑞先去加州大学的伯克利分校进修德文，秋季开学后又去阿拉巴马大学插班上四年级。他读书用功，学习成绩很优秀。1950年他获得化学学士学位后，随即进入宾夕法尼亚大学研究院的生物化学系，攻读导师威尔逊教授（D.W.Wilson）的博士学位，并于1955年获得了博士学位。接着他在Damon Runyon癌症研究基金会的资助下，来到美国东海岸的纽约市公共卫生研究所，开始进行博士后研究。短短几年内他已在有关领域发表了近20篇论文，并在博士后期满便成为该所的正式雇员。

1967年吴瑞和其领导的科研小组对DNA测序技术展开全面的研究，他利用天然存在的引物模板系统——大肠杆菌的λ噬菌体DNA的黏性末端作为引物，对黏性末端的DNA序列做了深入的研究，功夫不负有心人，他们历经3年多的潜心探索，于1970年在世界上首次成功地对λ噬菌体DNA的序列进行了解读，也成功解决了以前人们认为无法解决的技术难题。他们的研究成果发表在1971年5月的《分子生物学杂志》上，他的开创性工作也创立了能定位的引物延伸的方法，这促进了分子测序技术的发展。




DNA测序结果

吴瑞创建的能定位的引物延伸法进行DNA核苷酸顺序测定成功后，引起了科学界的重视。1973年桑格沿用这一方法的思路，改进了用聚丙烯酰胺凝胶电泳系统对标记的DNA进行分析的技术，于1975年建立了DNA测序的“加减法”，其中的“减法”主要是利用吴瑞的方法，1977年桑格又在“加减法”的基础上发展出“双脱氧法”，这种测序方法速度更快、更便利，并成为当今DNA分析的主要方法，因此他获得1980年的诺贝尔化学奖。不仅如此，穆利斯采用此引物延伸的方法后，于1985年发明了聚合酶链式反应（Polymerase Ohain Reaotion, PCR）技术，该技术可以在试管中快速获得数百万个特异DNA序列的拷贝，成为当今分子生物学中最有用的技术之一。因此，一些科学家称吴瑞教授为“诺贝尔级科学家”。

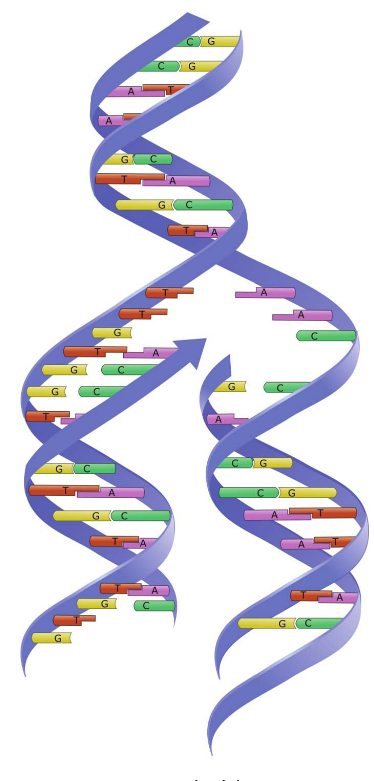
诺伯特·维纳（Norbert Wiener）在《控制论》中曾经有过这样的名言：到科学地图上的空白地区去做适当的勘察工作，只能由这样一群科学家来完成，他们每个人都是自己领域中的专家，并且对其邻近的领域都掌握了十分正确和熟练的知识。可以说DNA测序技术的建立就是一群科学家协同努力的结果，他们的工作相互联系，互相提供思路，互相借鉴，使得DNA测序技术的发展发生了重大的突破，大量的基因序列信息的解读，加快了分子生物学的发展。

## 7.4　盘旋的公路与PCR技术

近期面对新型冠状病毒奥密克戎毒株的肆虐，大家对于核酸检测技术已经不再陌生，但对于检测所使用的反转录聚合酶链式反应（reverse transcription PCR, RT-PCR）技术的诞生可能并没有太多的了解。

PCR技术是一种在生物体外用于扩增特定DNA片段的分子生物学技术，该技术自20世纪80年代问世后，在生物学、医学和法学等领域得到广泛应用，发挥了极其重要的作用。

印度裔美国科学家科拉纳早在20世纪50年代就已经合成了寡聚核苷酸。他利用体外合成的寡聚核苷酸合成酶以及DNA进行扩增，这一技术在当时被同行广泛使用，但是这项技术不能够严格地控制温度，对DNA聚合酶活性的影响强烈，所以仅能合成少量DNA,同时扩增率很低。根据自身多年的实验情况，科拉纳当时提出了两个重要的观点：一个是DNA暂且不能定序，另一个就是寡聚核苷酸体外合成相当困难。这种方法并没有明确地提出解聚再复合的观点，也并未弄清整个聚合的过程，直到1971年左右科拉纳又提出了核酸进行体外扩增的新想法，他认为可以通过DNA的变性，与合适的引物进行杂交，然后用DNA聚合酶来延伸引物，同时通过不停地循环进行该过程来扩增DNA。这种想法提出了一个大胆的假设，那就是在体外实现体内的生物学复制反应，但是当时尚未有相关的实验手段可以借鉴；同时更为重要的是具有较强稳定性的DNA聚合酶并没有被发现，因为在循环的过程中必须要升温到几十摄氏度的高温才能促成DNA的解聚，在这种温度条件下，非耐高温的DNA聚合酶都会变性失效，达不到聚合的目的；同时测序技术不成熟，合成适当的引物又相当困难，因此体外DNA的合成仍处在手工、半自动合成的阶段。所以这种思路仅仅是一种大胆的想法，并不能付诸实践。因为在实验中要经历几次高温，而在正常情况下，聚合酶的活性在高温下都将失活，不能继续下一次的反应，所以这个方案一直被搁置。



DNA复制

取得突破性进展的是美国生物学家凯利·穆利斯（Kary Mullis），1944年12月他出生在美国北卡罗来纳州南岭山附近一个偏僻的农村中，从小对生活中的事物充满着好奇心并且乐在其中。穆利斯的爸爸晚上带着穆利斯坐在厨房中，一边喝着啤酒一边告诉他当代加利福尼亚州的一些故事。父亲离世后，穆利斯还经常自斟自饮，思考问题。穆利斯的高中生活是在哥伦比亚度过的，在佐治亚工业学院化学系工作的时候他学到了很多有用的实验技术及数学、物理和化学知识。1966年穆利斯和妻子一起来到加州大学的伯克利分校，并在1972年获得加州大学伯克利分校生物化学博士学位。

1979年穆利斯进入西斯特公司，从事DNA的合成工作。穆利斯个性独特、不善合作，他在实验室里虽然与其他人常有矛盾，但是在生物实验方面的天赋还是得到大家的认可。1983年的4月到5月，穆利

斯一边在盘旋的公路上开车，一边在思考着如何解决这种体外复制难的问题，盘旋的公路和DNA双螺旋的相似性突然让他想到一种在体外复制DNA的方法模型，于是他开始进行资料的收集和整理，这是一块相对空白的领域，他决定并着手进行相关的实验。经过近3个月的准备，1983年8月穆利斯在西斯特公司做了关于PCR技术有关的学术报告，但是没有人能够相信，这种在体内复杂环境和相应的酶催化环境下进行的精密反应体系能够在体外实现复制，这虽然让穆利斯感到沮丧，但是他认为这种想法必须经过尝试。

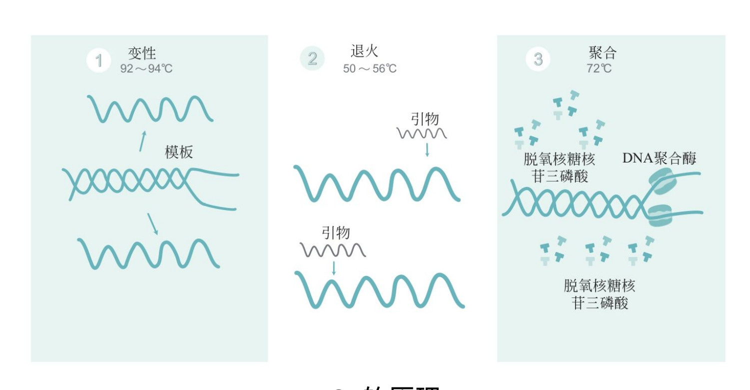
1983年9月，穆利斯和相关的几个实验员利用人体DNA做模板，抱着试试看的心态进行了世界上第一次的PCR实验，编号为PCR01，但是和很多人预想的一样，实验并没有成功。1983年10月他继续进行第二次实验，编号为PCR02，仍然没有成功。现在看来，实验的失败是多种因素造成的，包括选用的模板、实验室的温度、复制的环境、催化酶的活性等。1983年圣诞节前后，穆利斯又一次进行了PCR实验，改用了模板相对简单的pBR322质粒，随后又使用了噬菌体作为模板，实验结果有所改观但是仍不理想。1984年1月，穆利斯用自己合成的长寡聚核苷酸作为模板，扩增人的β珠蛋白基因的58个碱基对，实验终于取得了重大突破。西斯特公司决定让穆利斯成立独立的PCR实验研究小组，专门进行PCR技术的开发任务。

1984年11月15日，PCR实验终于获得了成功，1985年3月28日，西斯特公司申请了有关PCR的第一个专利。同年9月20日，一篇关于PCR技术应用的文章投稿到《科学》杂志，并在11月15日被接收发表。1986年5月，穆利斯应邀在冷泉港实验室举行的“人类分子生物学专题研讨会”上介绍了PCR技术，使得这项技术逐步走入公众视野。1991年霍夫曼-拉夫什公司出资3亿美元购买PCR技术，两年后穆利斯因为发明PCR技术而获得诺贝尔化学奖。



PCR仪

穆利斯设定的PCR条件包括温度、循环次数和扩增时间。温度与扩增时间是基于PCR反应的基本原理：变性到退火再到延伸的三个步骤而设置3个温度点。循环次数设置相对简单，但也是容易为人所忽略的一个步骤，标准的循环次数一般设置为25次，对于复杂模板如基因组DNA，循环次数必须进行摸索，有时增减一个循环次数就会导致整个实验结果的偏差。

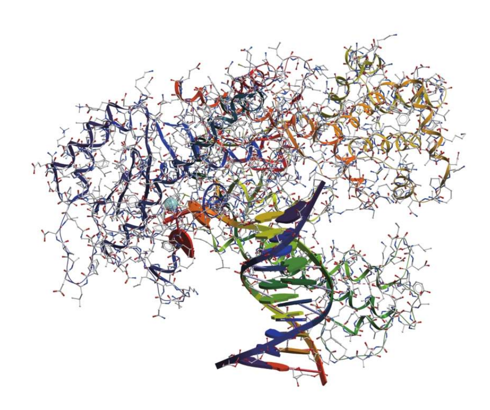



PCR的原理

## 7.5　温泉中的耐热聚合酶

PCR技术的发明是一个重大的突破，但是仍然不能被广泛地应用，最主要的原因是DNA聚合酶在多次的解聚、聚合的过程中活性变弱，同时聚合酶昂贵的价格也导致这种技术不能够广泛地普及使用。

实验伊始，因为实验的操作性复杂、成本高昂、可重复性不稳定等因素促使它成为“被搁浅的革新”，PCR技术改进研究也处于停滞的状态。其中最主要的原因是缺少可以耐高温的DNA聚合酶，每次实验一经过升温的步骤，DNA聚合酶就会失效导致下一步的扩增率非常低，甚至达不到扩增的效果。所以只能采取每循环一次便添加新的聚合酶的办法，这种半自动的扩增方法费时费力且效果并不明显，因为在每次新添加酶的时候并不能很好地控制实验的环境温度。

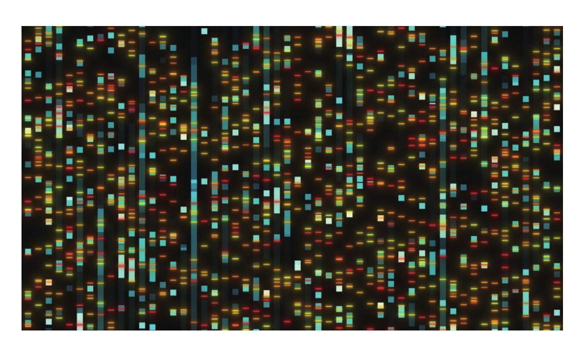


DNA聚合酶

直到1988年年初，美国科学家基奥哈诺（Keohanog）通过对所使用的酶的改进，显著提高了扩增的效果，使得扩增的循环次数逐步上升，这样扩增后的样本数量也大大增加。随后，日本科学家佐伯（Saiki）等从生活在温泉中的水生嗜热杆菌体内提取到一种耐热的DNA聚合酶，使得PCR技术的扩增效率大大提高。这种酶的发现，让DNA聚合酶在扩增的聚合步骤时不会失活，从而可以重复循环，达到提高自动扩增循环次数的功能，改变了原先半自动合成的现状，同时合成的次数也逐步上升。正是这种DNA聚合酶的发现使得PCR技术得到了广泛的应用，使该技术成为遗传与分子生物学分析的基石。

在PCR技术最终成熟之前，中国科学家钱嘉韵做出了不可磨灭的贡献。可以说她是将这种技术变为现实、踢出“临门一脚”的真正射手，她在研究中发现并成功地分离了耐高温的DNA聚合酶，而这种酶是PCR实现自动化的不可或缺的条件。钱嘉韵是出生在我国台湾省的

一名科学家，1973年她就读于美国俄亥俄州辛辛那提大学生物系，她的导师对黄石国家公园中温泉里的嗜热菌有着浓厚的研究兴趣，于是钱嘉韵在导师的建议下便开始了相关的学术研究。终于她不负众望，从这种嗜热菌中分离和提取了耐高温的DNA聚合酶，并于1976年在《细菌学杂志》上以第一作者的身份发表了相关的学术论文，得到科学界的广泛认可。很多公司的工作人员按照钱嘉韵论文中的实验操作步骤，也成功地分离了DNA聚合酶，使得PCR技术能够真正地实现自动、高效的复制。



基因DNA测序系谱分析可视化

1989年美国《科学》杂志将PCR技术列为10余项重大科学发明之首,并将1989年誉为“PCR爆炸年”。诺贝尔颁奖委员会认可了这项技术，并且在1993年将诺贝尔化学奖颁给了穆利斯，并且在颁奖词中宣称：PCR技术是一项分析DNA应用中最为广泛的技术，它可以将独立的遗传物质DNA片段在试管中复制数百万次，两个寡聚核苷酸通过正确结合能够成为可复制的物质，通过控制温度和酶形成新的寡聚核苷酸片段。通过不停地解聚、复合、解聚这样的循环往复的过程，在短短几小时的时间里可以让DNA的量增加上百万倍，这个过程是相当简单的，这一理论仅仅需要试管、热源及部分酶，即便现在商业上的PCR仪器需要很高的精确度，这一方法都是适用的。

总之，PCR技术的迅速发展拓宽了基础研究的前景，使基因克隆和测序变得简单起来。这项技术使得定点突变技术变得更加有效，同时PCR技术的发明，通过DNA的测序可以看清物种之间内在的进化顺序，对植物和动物分类学可能有重要的指导意义。伴随着测序技术和

PCR技术的日益发展，生物体DNA片段中蕴藏的大量信息被逐步解码出来，形成庞大的数据链，生物信息学在这种条件下应运而生，它是实验中衍生出来的一种行之有效的处理技术，它与测序技术相辅相成，互为补充，成为后基因组时代对生物体发育、代谢、疾病发生等一系列重大问题研究的突破口。

# 8 上帝之手——基因编辑

2020年10月7日，瑞典皇家科学院秘书长戈兰·汉森（Göran K. Hansson）教授宣布将诺贝尔化学奖授予美国化学家詹妮弗·杜德娜（Jennifer A.Doudna）和法国生物化学家埃马纽埃尔·卡彭蒂耶（Emmanuelle Charpentier），以奖励她们“开发了一种基因编辑的方法”。这项神秘的技术才逐步进入我们的视野。



基因编辑

什么是基因编辑技术呢？基因编辑技术是对生物体基因的特定位点进行精确修饰的技术，可以通过基因片段的替换、引入、敲除，达到改变生物体某一特性的作用。

## 8.1 给基因动手术

上面的说法可能听起来有点陌生，或者觉得替换、引入、敲除这样的字眼显得过于学术化了。下面我们先简单地介绍一下，为何要给基因动手术呢？在正常的状态下，我们体内的基因各司其职，各自负责生产不同的蛋白质来满足不同的生理要求和生活需求，但是有的时候，有些基因会发生一些突变，就仿佛基因生病了一样，生产出了不合格的蛋白质或者生产出来的蛋白质没有办法发挥自己的作用，因此就会引发一系列的疾病。这个时候，我们就需要对基因进行一次微观手术，对失效的部位进行切除、更换或者修补，就这么简单。

但是面对如此微小的结构，即使我们知道了操作的原理，但是实施起来依然犹如上天揽月般困难重重。

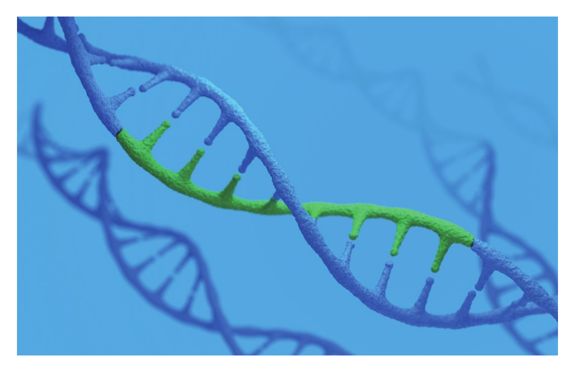


给基因动手术

## 8.2 找寻GPS、剪刀和针线

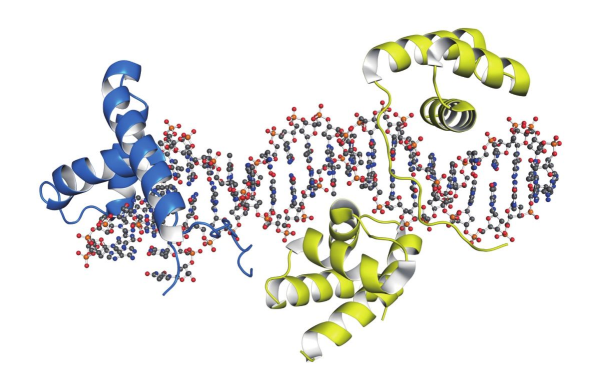
在了解完为何要进行基因编辑之后，下一步就是我们该如何来进行基因编辑。简单地说，我们要做这样的三件事：在活体细胞中找到一个特定的基因位点，再用剪刀将其剪下并把目标基因替换上去，最后用“针线”将这段基因和两边“邻居”基因位点牢牢缝补在一起。这样的过程听起来非常简单，但是寻找到合适的基因位点、用剪刀把目标基因裁剪下来、替换上新的基因进行缝合，哪一个步骤都难于登天。

要真正实现对基因的编辑，我们首先要找到要修改的目标基因组序列，这就需要一个精准的微小的GPS定位系统。1969年，美国生物化学家罗伯特·里德（Robert　Roeder）首次在真核生物中发现了三种不同功能的RNA聚合酶，它们是三种能以DNA为模板转录出RNA单链的蛋白质。1980年，里德实验室又发现了有些蛋白质能够帮助RNA聚合酶启动DNA转录过程，里德将它们称为“转录因子（transcription factor）”。其中一种转录因子TFIIIA（transcription　factor　IIIA）有一个不同于之前发现的转录因子的重要特征，它可以特异性地辅助RNA聚合酶启动基因的转录。简单来说，TFIIIA蛋白自带了GPS，可以在GPS的引导下精准到达目的地——基因位点，这一重大发现宣告精准找寻确切基因位点的GPS已经被发现了。



基因编辑后的DNA

接下来就是寻找可以将剪切好的DNA缝合起来的针线。20世纪90年代，美国纪念斯隆-凯特琳癌症中心（Memorial Sloan Kattering Cancer Center，MSKCC）的玛丽亚·雅辛（Maria Jasin）和同事在研究DNA断裂对肿瘤形成的影响时，发现“乳腺癌2号”基因发生突变会极大增加罹患乳腺癌和卵巢癌的风险，因为DNA断裂被修复的概率大大降低了。



与DNA结合的转录因子

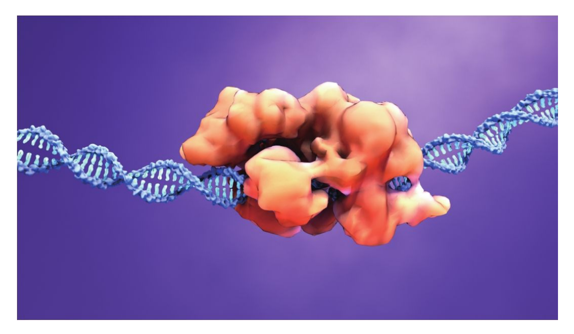
为了解决这个问题，就必须要找到可以自我修复的方法，就是看身体中有没有什么物质，就好比我们所说的“针线”，可以对这些断裂的DNA进行修复。接下来的研究她们发现了两种不同的“针线包”：一种叫作同源重组（homologous recombination，HR），以未受伤的姐妹染色单体的同源序列作为模板进行修复，虽然速度慢、效率低，但是由于严格依赖于DNA的同源性，可以使基因组修复到完好如初的效

果；另一种叫作非同源末端连接（non-homologous end joining, NHEJ），就是找到了两个DNA断点，为了防止DNA链的降解对生命体产生不可估量的损伤，就强行将两个DNA断裂处彼此连接在一起。这种修复方式简单粗暴却很高效，但是副作用明显，可能会造成序列缺失、片段插入、修复效果不精确等很多遗留问题。但是无论怎么说，这两种连接方式的发现，说明我们找寻已久的“针线”也已经浮出了水面。

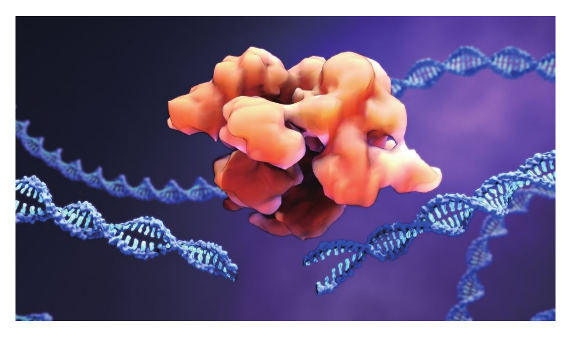
## 8.3 基因魔剪CRISPR/Cas9

1987年，日本科学家石野良纯（Yoshizumi Ishino）在分析大肠杆菌的DNA时，发现了一些奇妙的序列，这些序列很有意思，数十个碱基为一个单元的短序列重复了很多次形成一个长的序列，它的具体作用却并不明朗。

发现这样的序列纯属偶然，石野良纯当时在进行大肠杆菌DNA序列的分析，无意中发现了这个结构，当时的实验是为了研究其他的内容，但是出于好奇，石野良纯还是把这个重要的信息在文章中公布出来，为后人的研究埋下了伏笔。

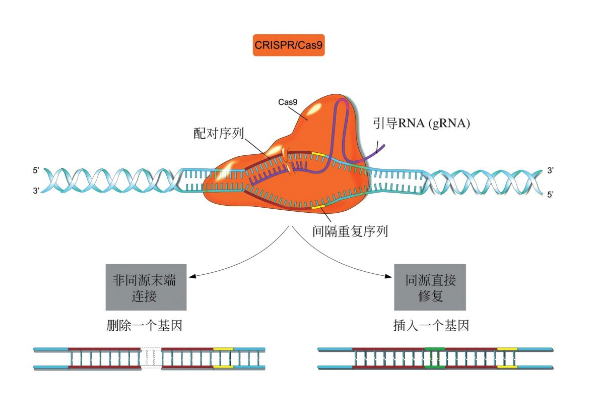


CRISPR/Cas9蛋白识别外来致病DNA



CRISPR/Cas9蛋白切割外来致病DNA

1993年，西班牙科学家弗朗西斯科·莫西卡（Francisco Mojica）在研究地中海嗜盐菌时，也在这些微生物的基因组里发现了这种奇怪的重复“回文”片段。这些片段长30个碱基，而且会不断重复。在两段重复之间，则是长约36个碱基的间隔。到2000年，莫西卡已经在20种不同细菌中发现了这种重复结构。但是他依旧对这样的结构有什么样的效用并不清楚，两年之后，这样的奇怪序列被发现的频率依旧在不停地增长，荷兰乌得勒支大学的吕德·扬森（Ruud Jansen）和同事给这些规律性重复的DNA片段起名为“成簇的规律间隔的短回文重复序列”（clustered regularly inter-spaced palindromic repeats），我们取这个名字的首字母，组成了“CRISPR”。



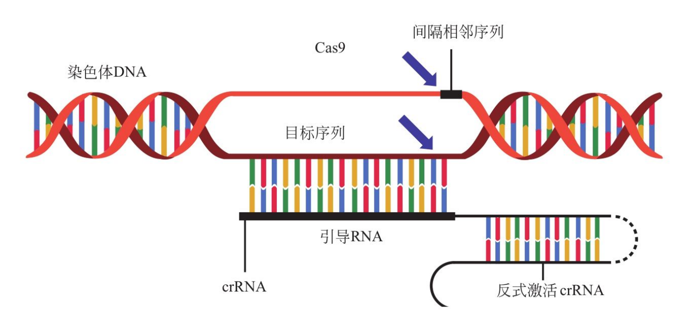

基因魔剪CRISPR/Cas9

同时，为了进行下一步研究，把临近CRISPR位点的基因命名为CRISPR-associated简称Cas。科研工作者发现Cas基因编码的蛋白质结

构与那些能与DNA发生相互作用的酶的结构很相似，似乎暗示着Cas基因与CRISPR序列之间存在某种功能上的相关性，但对于CRISPR系统的功能仍然一无所知。

紧接着，两位诺贝尔化学奖得主闪亮登场。加州大学伯克利分校的詹妮弗·杜德娜（Jenifer Doudna）在对细菌基因组序列进行测序的时候，发现了细菌中存在CRISPR序列。2011年3月，杜德娜参加了一场在波多黎各举办的细菌中RNA分子的研究会议，在会议中碰到了同样对CRISPR序列感兴趣的法国细菌生物学家埃马纽埃尔·卡彭蒂耶，两人一拍即合，准备合作进行研究。由于有着蛋白质结构研究和细菌学研究的基础，两人很快揭示了Cas9蛋白的工作原理：细菌通过CRISPR生成病毒DNA的RNA序列，再由RNA分子引导Cas9蛋白来到入侵病毒基因组的特定位点进行切割，造成病毒DNA双链的断裂。这一发现使她们意识到CRISPR/Cas序列兼具GPS和"剪刀"功能，它可以作为一种新的基因编辑工具。

2012年，两人将最新的研究成果发表在《科学》杂志上，宣告了最新一代基因编辑技术——CRISPR/Cas9的诞生。这一代技术相对于前两代技术有着得天独厚的优势：它不需要单独进行设计，只需要更换负责引导的RNA。换句话说就像我们要去一个地方，之前需要我们自己找寻交通工具，并且自己设计行进的路线，而现在的CRISPR/Cas9集成了汽车和手机的功能，我们只要在上面输入目的地，就可以让它自动行驶，载着我们到达目的地一样。



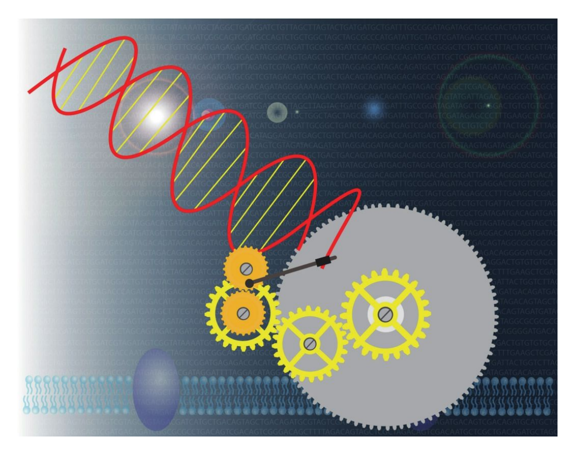

CRISPR/Cas9原理图

但是是否就说明这样的技术没有任何的缺陷呢？答案是否定的，目前CRISPR/Cas9系统也存在着容易“脱靶”的缺点。由于这个系统是一种单链，自身稳定性差，容易发生突变，所以会有一定概率的“脱靶”现象发生，给实验带来不确定性。但是不管怎么说，目前CRISPR/Cas9依旧是最行之有效的DNA编辑工具。

## 8.4　潘多拉的魔盒

既然我们已经掌握了这种可以修改基因的工具，那么是不是就意味着我们已经触摸到了我们生命中最核心的本质，也开始掌控住了人类命运主宰的上帝之手。但是这一定是一件好事吗？

其实很显然，任何一项技术没有好坏之分，而是看使用它的人和如何去使用它，这是一把“双刃剑”。



基因工程

转基因技术是利用现代生物技术，将人们期望的目标基因经过人工分离、重组后，导入并整合到生物体的基因组中，从而改变生物原有的性状或赋予其新的性状。自从“基因”的概念深入人心之后，“转基因”这一词汇也逐渐融入人们的生活中。随着生物技术的发展，人类不禁会思考，我们是不是可以充当“上帝”的角色，自己去创造新的物种。

一方面，基因编辑可以通过它独特的方式治愈部分遗传疾病，但是另一方面，基因编辑后的基因可能带有明显的可遗传性，它会对我们现有的人类基因库造成毁灭性的打击。2017年2月，美国国家科学院与医学院联合公布了一份名为《人类基因编辑：科学、伦理以及监管》的报告，在这篇报告中，为基因编辑的研究设置了一条红线：有关人类胚胎细胞的基因编辑基础研究不会受到限制。但是，如果将其作为临床应用，医治患者，则需要经过非常严格的伦理审批才可以进行，其使用范围也非常有限，一般仅限于对严重疾病使用，并且需要在患者对病情有充分认识的情况下使用。



转基因工程

目前关于基因编辑的研究主要集中在进行相关疾病的基因改造上，而我们说对于非疾病类的基因编辑应该给予更多的关注和评估，因为这带来的后果可能是现在无法估量的。

之前也存在一些惨痛的案例。早在2000年法国内克尔医院的研究组就利用基因编辑技术对X连锁重症联合免疫缺陷症进行治疗，当时治疗大获成功并且获得了一片喝彩，但是后续的发展却不尽如人意。没过多久，20名接受治疗的患者中，有5名发展成为了白血病患者，并且有1人死亡。后续的调查发现，在进行基因编辑的过程中随机插

入了可以激活癌症基因表达的基因序列。随后的研究中，有科学家发现，接受基因编辑后的短时间内，也许不会产生明显的副作用，但是从长期来看，有可能会导致一些不可逆的损伤。例如，针对肌肉组织进行的基因编辑，可以强化肌肉的纤维组织，可能会在1～2年内导致多巴胺分泌异常，损坏人体机能，增加阿尔茨海默症的发病率，存在一定的安全隐患。

人类的发展与科技息息相关，科技的进步在不断地改变着人类的生活，也在悄无声息地改变着人类自己。很多人认为人类可以主宰“自然”，成为自己的“造物主”。然而，无数的事实证明，这是非常荒诞和愚蠢的！大自然是神奇的，生命是多么的不可思议，是多么的值得敬畏！人类必须对生命怀有一颗敬畏之心，不忘初心，方得始终！令人欣慰的是，许多国家已经赋予了生物伦理学“法”的地位，并针对各自的国情，采取了不同的应对措施。

生物伦理学已经在规范人类的社会行为方面起到了日益重要的作用，因此应结合科技发展，不断地调整生物伦理学的关注视角，使它始终向着符合社会规范的方向发展。荀子说：“水火有气而无生，草木有生而无知，禽兽有知而无义，人有气，有生，有知，亦且有义，故为天下最贵也。”只要人类固守生物伦理学的底线，就一定能够促进全人类的可持续发展。



# 9　破解感知之谜——温度与痛觉

医院里，注射器的针尖闪着寒光猛地扎进肌肉，短暂的停留之后，带来一阵锥心的疼痛。和恋人手牵手漫步沙滩，有时候能够感受到彼此指尖那触电般的甜蜜感觉。面对寒冷的冬日，或者炎炎的夏日，我们能够清楚地感受到温度带来的冷热刺激。感受疼痛，感知自我，感通世界，是我们最习以为常的能力。然而，这种能力究竟从何而来，又是如何发挥作用，自古以来一直是困扰人类的谜题。




温度感知

## 9.1　感知外界的受体

我们知道人类主要有5种感觉，包括视觉、听觉、嗅觉、味觉和触觉，此外人类还具有感受冷热刺激的温觉及痛觉等。但是人类是如何感知这些物理性的变化，自古以来一直是生物学层面难以解释的难题。

也许大家有这样的感觉，我们刚穿上衣服的时候会感觉到衣服的存在，刚穿上袜子的时候也能感觉到袜子的存在，刚戴上手表的时候能感觉到手表的存在，可是过了一段时间之后，这种感觉就会逐渐消失。说明我们有感知外界的能力，也有控制这种感知的能力，否则长时间无用的刺激会干扰到我们的注意力，而拥有这种控制感受的能力

才是必须的。

感知冷、热和触觉的能力对于人类生存至关重要。在古代中国，也有对触觉简单的描述，《易传》中提到“寒暑相推而岁成焉”，《荀子》中提到“温润而泽，仁也”，《孟子》中说“文王视民如伤，以痛为爱”。在古代西方，亚里士多德也系统地描述了触觉感知，他把触觉置于经验感官的中心，认为触觉很可能是各种感觉能力的综合，是最基本最重要的感觉。

17世纪哲学家勒内·笛卡儿（René Descartes）设想出将皮肤的不同部分与大脑连接起来的线，通过这种方式，人们接触明火的脚会向大脑发送相应的机械信号，这可能是对于感知的最初的想法。18世纪，法国哲学家埃蒂耶纳·博诺·德·孔狄亚克（Etienne Bonnot de Condillac）论述了触觉认知对人类智能的重要性，他认为触觉是感官世界的中心，其他的感官都是以触觉为基础复合而成的。



大脑认知

1944年，约瑟夫·厄兰格（Joseph Erlanger）和赫伯特·加瑟（Herbert Gasser）合作，将阴极射线示波器用于对神经动作电位的研究。他们由于发现了对不同刺激做出反应的不同类型的感觉神经纤维而获得了诺贝尔生理学或医学奖。从此以后，触觉的神经生理学机制研究开始活跃起来，研究表明，触觉感知到的特征信息被转化成动作电位传递到大脑中，大脑的高级联合皮层对外界刺激进行精细分析，结合过去的经验，就会产生触觉认知。

之后的研究依旧继续进行，但是始终没有掀起太大的波澜，直到瑞典当地时间2021年10月4日，瑞典皇家科学院宣布将2021年诺贝尔生理学或医学奖颁发给美国加利福尼亚大学的生理学家戴维·朱利叶斯（David Julies）教授与加利福尼亚州拉霍亚斯克利普斯研究所（Scripps Research）的亚美尼亚裔美籍神经科学家阿尔登·帕塔普蒂安（Ardem Patapoutian）教授，以表彰他们发现了温度和触觉的受体。我们这才对这种司空见惯的神奇过程有了更深层次的了解。




戴维·朱利叶斯（左）和阿尔登·帕塔普蒂安（右）

两位科学家发现了某些基因具有一些特殊的功能，可以加深我们对冷热触发皮肤感知的理解，也能够让我们知晓机械性的外力触发人体神经冲动的机制。其中，具有特殊功能的受体包括：辣椒素受体TRPV1（transient receptor potential vanilloid type 1）、冷和薄荷醇受体TRPM8（transient receptor potential cation channel subfamily M member 8）和Piezo（希腊语“压力”），这些基因的突破性发现和对它们的研究过程，让我们知晓了人体内多种生理过程的机制，也能够有针对性地对慢性疼痛、呼吸系统疾病、神经系统疾病等提出诊疗的方案。

诺贝尔颁奖委员会称：“两位获奖者指出，在我们理解感官与环境之间复杂的相互作用时，存在着关键的缺失环节。他们的发现通过解释感知冷热和机械力的分子基础，从而解开了大自然的一个秘密，这是我们感知及与内外环境互动的基础。这些突破性的成果推动了感知领域的研究浪潮，使得我们对神经系统如何感知冷热和机械刺激的

理解迅速增加。”

## 9.2　人类的礼物——痛和痒

痛觉是动物在受到伤害时的一种“警告”，可引发机体一系列防御性的保护反应。为什么会演化出这种能力，是因为相比较植物而言，动物更经不起外来刺激对身体的伤害。

植物相对来说构造简单，身体中的各种部位没有固定的形状，甚至失去身体中的绝大部分组织，都不会危及植物的生命。与植物相比，动物构造要复杂许多，拥有各种不同的循环系统：淋巴系统、血液系统、骨骼系统、泌尿系统等，如果身体的某一个微小的部分受到伤害，很可能会危及生命安全。因此，我们必须演化出一种能够应对这种伤害的感应机制，那就是痛觉，让我们通过痛觉感受到危险存在，对危险做出激烈的反应，从而在以后的生活中规避这些伤害。

在医学上，疼痛是最常见的症状之一，疼痛发生的位置通常指示病灶所在，疼痛的性质也间接说明了病理过程的类型。



颈椎疼痛

另外，除了痛觉之外，还有一种感觉，那就是痒。痒是一种动物身体对轻微的、潜在有害刺激的反应。这种感觉带来的伤害不会危及动物的生命，但是让我们有一种酥麻的感觉。如蚊虫的叮咬、皮肤的瘙痒、昆虫从皮肤爬过等，这种感觉可以提醒身体注意，有不良刺激存在，但是危险性不大。因此不需要逃离这些危险，只需要用手去抓挠即可。

此外，伤口愈合会感觉到痒、胆道阻塞会感觉到痒、镇痛的吗啡也会引起痒，身体中有多种物质能够让我们感受到痒的存在。

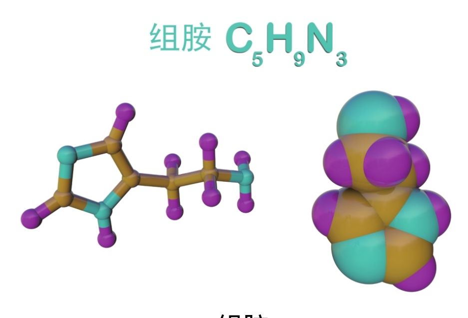
痒和痛是类似的，也是皮肤感受到的一种不愉快的感觉，都是通过脊髓–丘脑传递到大脑的感觉传导通路，因此很多人把痒的这种感觉认为是“微痛”，但是这是错误的。两者之间有联系，但是又完全不同。痛可以来自于皮肤，也可以来源于内脏、关节、肌肉，但是痒只能来自于皮肤和接近体表的黏膜，如果痒只是微痛的话，那么肌肉和关节应该能感觉到痒才对，但是在生活中我们的肌肉、关节等部位是无法感受到痒的，说明这种说法是不正确的。

许多慢性瘙痒症的患者感受到的痒对身体没有任何好处，反而会影响到患者的生活质量，那么为什么会演化出痒这种感觉呢？

我们知道，当皮肤受到刺激的时候会产生组胺，组胺是一种可以导致炎症的物质，可以使得皮肤红肿，让人感觉到痒。在正常皮肤的浅表处注射组胺也能引发痒的感觉；注射血清素也能让动物引起痒的感觉；胆酸与受体结合能够通过G蛋白使细胞内的钙离子浓度升高，活化神经细胞，让人产生痒的感觉……

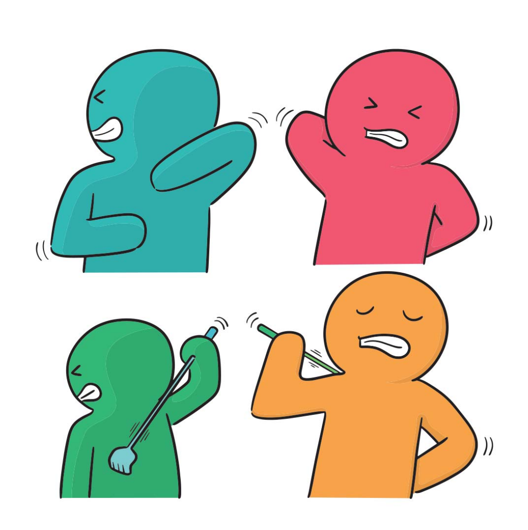


痒




组胺

在脊椎动物体内引发痒的组胺在低等的线虫身上并不存在，线虫已经发育出了触觉和痛觉，但是没有痒的感觉；在果蝇、蝗虫、蜜蜂等昆虫体内，已经存在了组胺和组胺的受体，可是这些受体在传入神经中并不存在，因此昆虫还是依赖触觉和痛觉来代替痒的感觉；鱼类也没有痒的感觉，只有四足动物才有，因为它们具备了四肢，可以抓挠身体的各个部分。



抓挠去除瘙痒

我们可以大胆地猜测，在生物演化的过程中，仅仅具有触觉和痛觉是不够的，需要更为精确的感觉来区分各种刺激。如无害的触觉、有害的痛觉、可能会引起局部伤害但是还不会致命的感觉等，因此演化出了痒的感觉，让我们通过抓挠的动作去驱逐刺激源，从而更精确

地对于外界的接触做出反应。

## 9.3 神奇的辣椒素受体

提到外界的刺激，辣是一种让人又爱又恨的感觉。早在公元前5000年，玛雅人就开始种植和食用辣椒，辣椒也是人类种植的最为古老的农作物之一。在吃辣椒的时候，我们随着辣度的提升，会感觉到明显的灼热甚至疼痛，这是因为辣椒中含有的辣椒素可以激活人体伤害性初级感觉神经元。但是辣椒究竟如何将这种物理性刺激转化为生物性刺激的过程依旧不得而知。



小米辣

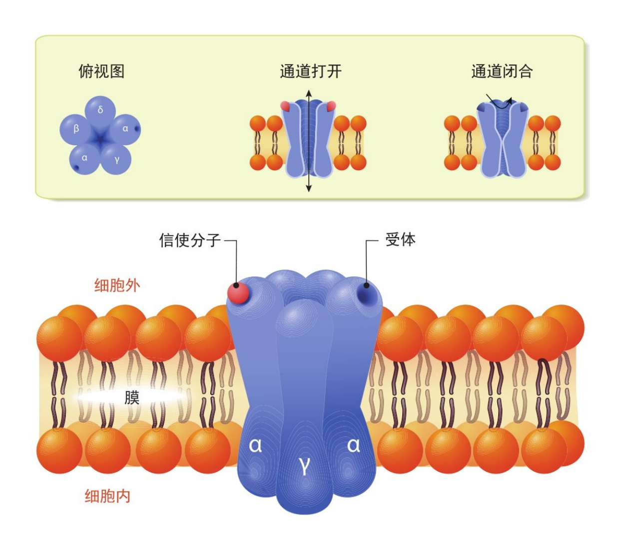
最先在这个领域做出杰出贡献的就是之前提到的戴维·朱利叶斯。1955年，朱利叶斯生于美国纽约布鲁克林区，他在布莱顿海滩边长大，他的祖父母20世纪初随着东欧移民潮来到美国，朱利叶斯的父亲是一名电气工程师，母亲是一名小学教师，重视教育的家族传统让他从小就知道持续学习和接受高等教育是人生计划的一部分。

1973年，朱利叶斯考入麻省理工学院。大三时，他结识了生物大分子研究领域的杰出人物之一亚历山大·里奇（Alexander Rich），并加入了他的实验室。 1977年大学毕业后，他进入加州大学伯克利分校杰里米·索纳（Jeremy Thorner）和兰迪·舍克曼（Randy Schekman）的实验室攻读硕士研究生。1989年，朱利叶斯开始在加州大学旧金山分校担任助理教授，他的科研小组继续研究神经药理学

和神经生理学上的基本问题。功夫不负有心人，伴随着在神经科学领域的逐步积累，他开始追问自己“疼痛”在分子生理学上的具体成因，而正是这一问题的提出，指引着朱利叶斯来到“感知”谜题的山脚下。

我们在吃辣椒的时候，能够感受到一种火热的灼热感。辣椒中的辣椒素是如何引起我们的神经发生反应、激活引起疼痛的神经细胞是我们一直想弄明白的问题。

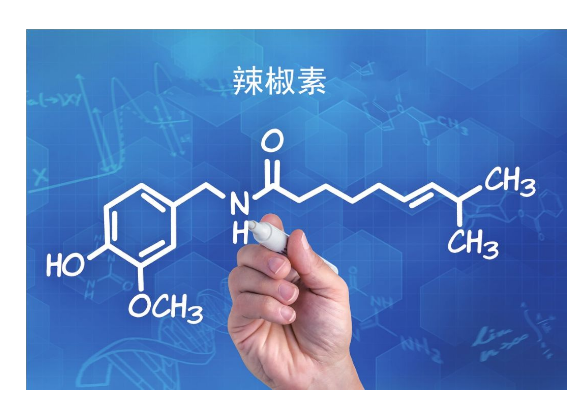
朱利叶斯和同事进行了一系列的筛选工作，他们发现对辣椒素敏感的痛觉感受器主要分布在感受器神经节，所以他们构建了一个背根神经节的cDNA文库，包含了大约16000个克隆。然后把这些克隆分为若干实验组，分别转染到不表达辣椒素受体的HEK293细胞中，同时加入钙离子荧光探针Fura-2，通过钙成像来观测辣椒素诱导的细胞内钙离子浓度变化。如果某一组细胞内的钙离子浓度升高，就说明该组克隆中可能含有辣椒素受体。




离子通道蛋白

经过了长时间的表达筛选，朱利叶斯最终鉴定出一种对辣椒素敏感的新的基因，这个基因编码的辣椒素受体是一个新的离子通道蛋白，这种离子通道蛋白就被命名为TRPV1，又称为VR1（vanilloid receptor subtype1）。

之前朱利叶斯的研究发现，离子通道蛋白TRPV1还和痛觉有着密切的联系，尤其是与炎症相关的疼痛。那么在医学上就给缓解疼痛开辟了一个新的研究方向。如果我们找到了可以抑制TRPV1发挥作用的药物，就能够降低人体对疼痛的感觉，这样就达到了缓解疼痛的作用，对饱受慢性疼痛折磨的病人来说是个好消息，也可以在缓解癌症疼痛上提出新的治疗方案。




辣椒素化学结构式

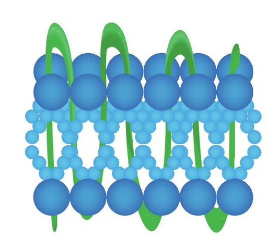
沿着这样的思路，随后很多科学家在人体中发现了更多的TRP类受体，如会被芥末激活的TRPA1受体、会被百里香激活的TRPA3受体。相关研究还发现，在一些癌症患者体内，多种TRP受体出现了一定程度的变化，这个研究成果表明，我们可以通过TRP受体来提早发现和治疗癌症。

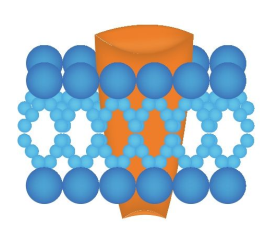
## 9.4　触觉的受体

与朱利叶斯共同获得诺贝尔生理学或医学奖的还有阿尔登·帕塔普蒂安（Ardem Patapoutian）。1967年他出生于黎巴嫩贝鲁特，1986年跟随父母搬迁到洛杉矶。帕塔普蒂安在回忆时说道："也许和我小时候在贝鲁特所经历的一样，在洛杉矶的第一年也是一场不同寻常的适应斗争。"为了在当地生存，他做了一年"不拘一格"的工作，如贩卖比萨饼和为亚美尼亚报纸撰写每周星座运势的文章。坎坷的成长历程培养了他坚忍不拔的意志和极强的适应能力，这在他以后的科研工作

中发挥了重要作用。

1986年，帕塔普蒂安考入加州大学洛杉矶分校，大学期间他加入了朱迪思·安·伦吉尔（Judith Ann Lengyel）教授的实验室，并在理查德·巴尔达雷利（Richard Baldarelli）教授的指导下学习分子生物学。1990年大学毕业后，帕塔普蒂安继续发育生物学转录调控的研究。1996年，他加入加州大学旧金山分校路易斯·雷查德（Louis Reichardt）教授的实验室做博士后研究员，对引发触觉和疼痛体感神经元的发育程序展开研究。他在回忆中说道："因为这些感知系统仍然如此神秘。当你发现一个不太了解的领域时，就是一个深入挖掘的好机会。"





G蛋白和离子通道

在研究感觉神经元发育的时间里，帕塔普蒂安逐渐意识到，作为这些细胞功能基础的蛋白质——使它们能够检测温度和机械力等物理刺激的生物大分子，目前还处于一种未知状态。究竟是哪些通道将机械力转化为神经元信号，从而启动触觉、本体感觉和疼痛呢？

TRPV1离子通道蛋白的发现是一项重大突破，为发现其他的温度感应受体开辟了新道路。帕塔普蒂安和朱利叶斯分别独立地使用薄荷醇鉴定出被冷激活的温度受体TRPM8。这些离子通道蛋白的作用揭秘，表明它们可以被不同的温度激活，用基因缺失小鼠的实验也证明了这些通道蛋白的作用，也回答了肌体如何对温度做出应答的。当大部分温度受体被发现之后，帕塔普蒂安决定向感觉神经受体领域的制高点——触觉受体发起挑战。

帕塔普蒂安和同事首先鉴定出一种可以在实验室培养皿中生长的胶质瘤细胞系，当用微量移液管触碰该细胞系中的单个细胞时，该细

胞会发出可测量的电信号，可以被专门设计的细胞探针观测到。他们推测这样一种可以被机械力激活的受体应该是一种离子蛋白通道。

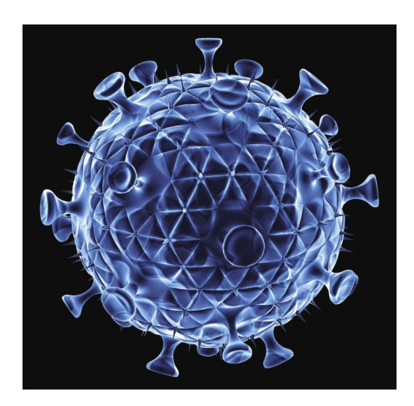
随后帕塔普蒂安从人类2万多个编码基因中挑选出300多个可以在实验中高表达的基因，再从中挑选出72个可能编码受体的候选基因进行筛选，用分子生物学的方法依次将其敲除，然后再测量失去了某个基因的小鼠细胞对压力的反应。尽管上述实验设计得十分巧妙，帕塔普蒂安及其同事们仍然花费了3年多的时间才鉴定出第一个基因，这个基因的失活会导致细胞失去对触碰机械刺激的敏感性。因此，这个基因被命名为Piezo1，这个词来源于希腊语中的“压力”一词。

随后，帕塔普蒂安在工作中又发现了第二个触觉基因，并将其命名为Piezo2。进一步的研究表明Piezo1和Piezo2分别参与调控肌体的很多重要生理功能，包括血压调节、呼吸、膀胱控制等。

简单来说，TRPV1、TRPM8、Piezo离子通道蛋白的发现，让人类了解到，我们是如何感知和适应外界的环境变化，TRP通道是肌体感知温度的关键，而Piezo通道则赋予了我们感应触觉刺激的能力。除此之外，这些离子通道蛋白还在许多与感受温度和机械刺激有关的重要生理过程中起关键作用。至此，我们才逐步打开人体感知的大门，了解到外界刺激如何为我们所感知，但是未来还有更多更漫长的研究之路等待着我们去探索。

# 10 人类生命的黑板擦——埃博拉病毒

在历史上，人类曾经遭受过多种病毒的袭击，包括天花病毒、狂犬病毒、禽流感病毒、埃博拉病毒、非典型肺炎病毒……随着科技的进步、交通的便捷、抗生素的滥用、人类对大自然的破坏、人类对野生动物的捕食等，病毒性传染病的暴发频率越来越高，出现了很多毒性和传染性更强的、我们从未见过的新型病毒。这些病毒引发的传染病给人类带来了巨大的灾难。



病毒粒子

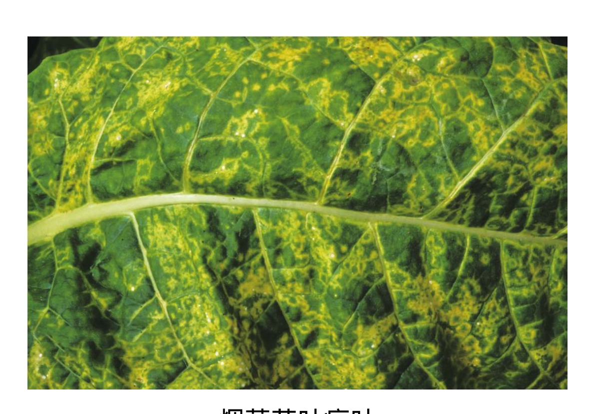
## 10.1 病毒的发现

在我们了解埃博拉病毒之前，先要了解什么是病毒，病毒又是如何被命名的。

根据微生物学中的病毒定义，病毒是一类由核酸和蛋白质等少数几种成分组成的超显微的非细胞生物，只含有DNA或RNA的遗传因子，能以感染态和非感染态两种状态存在。病毒广泛寄生于动物、植物、微生物细胞中，是一种细胞内专性寄生的特殊生命形式，具有高度侵染性，无细胞结构，体积非常微小，是具有基因、可复制和进化并占有一定生态地位的生物实体。

既然病毒这么小，那么它是如何被发现和命名的呢？第一位我们需要了解的人物就是在荷兰工作的德国农艺化学家阿道夫·爱德华·麦尔（Adolf Eduard Mayer），1886年他首次发现并且命名了烟草花叶病毒。麦尔出生于德国西北部的奥尔登堡，17岁考入德国卡尔斯鲁厄理工学院学习数学和化学，1862年继续进入德国海德堡大学攻读博士学位，毕业之后留校担任了发酵学和化学工程两门课程的主讲教师。1876年，麦尔前往荷兰担任瓦格宁根农业试验站的主任，从此与烟草花叶病毒有了不解之缘。

19世纪末，烟草行业发展迅速，成为很多国家的支柱产业。但是好景不长，很多地区的种植户发现，他们种植的烟草叶得了一种奇怪的病，患病的叶片上会出现坏死斑和黄绿相间的条纹，随后叶片出现肥厚不一、营养不良等状况，烟叶就不能再继续使用了，这给种植户带来了极大的损失。



烟草花叶病叶

由于找不到烟草花叶病的发病原因，种植户也没有针对性的方法，只能放任这些烟草叶在田地中逐渐腐烂。1879年，麦尔开始对这种烟草花叶病进行细致的研究。他开始从温度、光照、种子等角度对健康的烟叶和患病的烟叶进行细致的比较，结果并没有发现有什么异样，也就是说，这些条件对于烟草花叶病的发生没有任何的影响。当这些因素都排除之后，他认为可能是土壤在作祟，可能由于土壤中缺少了某些微量元素或者含有某些重金属元素才导致这种斑纹的产生。



种植烟草叶

经过对土壤成分的化验，结果让麦尔大失所望，健康植株和患病植株生活的土壤成分基本没有区别。此时，麦尔想到了法国微生物学家巴斯德和科赫的实验方法，将患病的烟草叶研磨后放在显微镜下观察，是否能够发现致病因子呢？他先把叶片放入器皿中研磨，得到带着叶肉组织的汁水，然后在显微镜下观察这些汁水的涂片，希望能够发现细小的微生物。但遗憾的是，他什么微生物也没观察到。虽然在显微镜下没能发现任何的蛛丝马迹，但是人们却发现，如果把磨碎的患病叶片的汁液放在正常的叶片上，原本健康的叶片很快就会染病，这说明患病叶片的汁液中存在致病的微生物。如果把这种汁液稀释100万倍，然后将稀释后的液体涂在健康的叶片上，原本健康的叶片仍然会染病，说明此种微生物的生命力是极其顽强的。如果将从患有烟草花叶病的烟草叶中提取的汁液加热到80摄氏度之后，再进行感染，结果发现，汁液失去了感染的能力。一系列的实验让人们确信，病叶上一定存在着一种致病的因子，只是一时无法观察到它。于是，麦尔在1882年把这种烟草疾病命名为“烟草花叶病”（tobacco mosaic disease）。

随后，俄国生物学家伊凡诺夫斯基（Ivanovsky）发现了烟草花叶病的致病因子具有滤过性。1888年，年仅22岁的伊凡诺夫斯基获得了圣彼得堡大学的理学学士学位，他研究的课题是“论烟草的两种疾病”。在大学期间，他就参加了俄罗斯农业部的研究课题，主要是为了探明烟草花叶病的成因。毕业之后，他再度研究烟草花叶病的成

因，他与麦尔可以说是在同一时间从不同的角度进行着病因的探索。

伊凡诺夫斯基使用当时最先进的生产无菌纯净水的过滤器——尚柏朗氏过滤器进行实验，过滤患病烟草花叶的汁液，但是发现其依然具有传染性，伊凡诺夫斯基认为烟草花叶病毒的致病因子也是一种细菌，只不过尺寸更小而已。

之后，真正对病毒进行命名的是荷兰细菌学教授贝杰林克（M.Beijenrinck）。1851年贝杰林克出生于阿姆斯特丹，1872年从代尔夫特理工学院化学系毕业，并考入莱顿大学，在1877年获得了理学博士学位。1876年贝杰林克和麦尔一起来到瓦格宁根农学院担任植物学教师。麦尔比贝杰林克年长几岁，贝杰林克就经常向麦尔请教烟草花叶病方面的问题，并且获得了很多有价值的信息。1885年，贝杰林克离开了瓦格宁根，但是关于烟草花叶病的研究却始终没有停止。



烟草花叶病毒

1895年，贝杰林克来到代尔夫特理工学院担任细菌学教授，两年后新的细菌学实验室和温室建成。贝杰林克重新开始了烟草花叶病毒的研究，他利用尚柏朗氏过滤器进行过滤，结果发现，滤液依旧具有感染性。随后他进行了一系列的对比试验，首先将滤液稀释成不同的比例，分别进行健康叶片的感染，试验结果表明，不同浓度的汁液都无差别地具有相同的感染能力，说明这肯定不是由无生命的化学物质，如毒素引起的；第二个对比试验是分别用蒸馏水和健康叶片的汁

液进行稀释，得到的液体再感染健康叶片，结果发现二者的感染程度相同，说明滤液并没有在健康叶片的汁液中发生增殖；贝杰林克在此基础上继续做了琼脂糖凝胶扩散实验，结果发现滤液可以发生扩散，并且扩散后的滤液依旧具有传染性，说明烟草花叶病的致病因子是一种液体或者是可溶的。

1898年，贝杰林克发表了关于自己实验结果的论文，提出了“传染性活流质”（contagium vivumfluidum）的概念。在论文中，贝杰林克都是用contagium（触染物）和virus（病毒）的概念进行阐述。贝杰林克的病毒概念，主要包括以下几点内容：一是可以通过细菌滤器，二是可以具有传染性，三是能在生物体内增殖，但是不能在体外生长。其实，virus这个词最早是一世纪由罗马的名医塞尔萨斯提出的，他当时对于virus的拉丁文定义是“黏液”，与现在的病毒含义是大相径庭的。贝杰林克沿用了塞尔萨斯提出的virus的原有的“黏液”的含义，并且在此基础上继续发展出了病毒的概念。

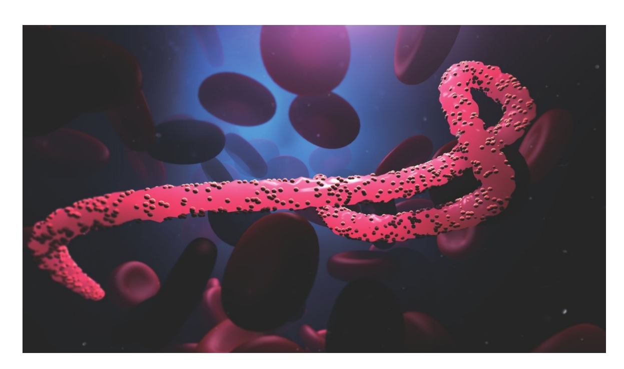
## 10.2 埃博拉病毒的暴发

病毒作为比人类更加古老的一类特殊的物种存在，它有着很多独有的特点。而在这些病毒中，有一种令人闻风丧胆的病毒，那就是被称为人类生命黑板擦的埃博拉病毒。

1976年，在苏丹南部和刚果（金）（旧称扎伊尔）的埃博拉河附近发现了一种急性出血性传染病，最早的病例可以溯源到一个名叫亚布库的小村子，村子里有一所学校，学校的校长从扎伊尔北部旅游回来后就感觉到身体不适，但是他没有特别在意。1976年8月26日，这位校长，也是这次感染的零号病人开始发烧，但是当地的医院查不出来病因，就提供了一些奎宁让他回家自行调养。“奎宁”在秘鲁文字中是树皮的意思，对疟疾等病有良好的疗效。回到家中调养一段时间后，他发现病症并没有减轻，10天之后的9月5日，校长因为病情加重不得不再次前往当地的一所教会医院就诊，随后他的病情加速恶化，最终在9月8日不治去世。

之后连锁反应逐步出现，校长去世后不久，在家中照顾他的亲

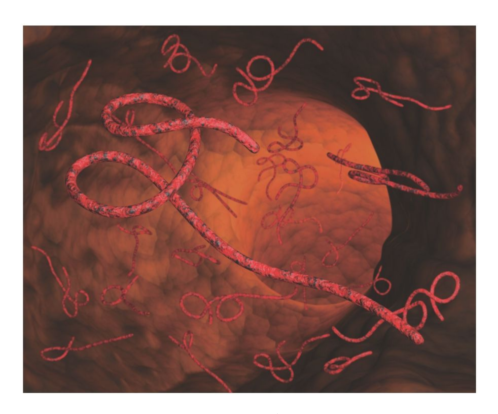
人，以及教会医院里的修女，陆续出现发病的现象。病人在发烧的同时，出现浑身疼痛、剧烈呕吐和腹泻，并且伴有七窍和内脏出血，最后在短时间内死亡。



带有血细胞的埃博拉病毒

这种可怕的疾病立刻引起了人们的重视，在一位修女感染疾病去世后，她的血液样本被送到了比利时的病毒研究所。当时27岁的研究人员彼得·皮奥特（Peter　Piot）在显微镜下看到了一种丝状的病毒，这种病毒是他之前从未见过的病毒类型，因此他怀着一颗好奇心，坚持前往扎伊尔进行新型病毒的研究和疫情防控。皮奥特来到扎伊尔之后发现，参加葬礼的病逝者亲友和遗体会有亲密接触，这可能是一个潜在的传染风险。因此，皮奥特坚持妥善处理遗体，禁止直系亲属直接接触遗体，并且采取了相应的隔离和保护措施，最终控制住疫情的蔓延。这场疫情共造成了318人患病，280人死亡，致死率接近90%，远远超过天花、霍乱、鼠疫等传染病，令人谈之色变。在病源地的村庄旁有一条埃博拉河，当地村民们认为这是上帝对他们的惩罚，因此皮奥特将该病毒命名为埃博拉病毒（Ebola Virus, EBoV）。

埃博拉出血热是由埃博拉病毒导致的血管破裂而出现的全身性的急性出血热传染病，临床上主要表现为急性发病、发热、呕吐、肌肉疼痛、出血、肝肾功能受损。埃博拉病毒感染者会出现严重的出血现象，内脏器官会逐步地发生类似于融化的现象，并导致休克综合征，症状非常的恐怖。



埃博拉病毒

由于埃博拉病毒的高致死率，再加上初次暴发地地处偏远的小村落，交通闭塞，人员流动并不频繁，首次暴发之后很快便销声匿迹了。

随着交通的便捷，人员流动的频繁，埃博拉病毒再次如幽灵般地闪现。2012年年底，非洲几内亚的一名2岁小男孩感染了埃博拉病毒，这个小男孩很快便去世了。紧接着，“不祥”的事情接二连三地发生，小男孩的姐姐和母亲先后感染，并且将病毒传染给了附近村庄的人。由于首次发病的零号病人——小男孩居住的村庄位于几内亚、利比里亚、塞拉利昂三国的交界处，所以导致了病毒在这三个国家快速传播。2014年2月，埃博拉病毒在几内亚境内暴发，并且波及利比里亚、塞拉利昂、尼日利亚、塞内加尔、美国、西班牙、马里等七个国家。据世界卫生组织统计，截至2014年11月15日，全球共计有15145人感染埃博拉病毒，死亡5420人。2018年8月1日，刚果（金）暴发了埃博拉病毒发现以来的最大规模的疫情，截至2020年2月该国的确诊感染人数达3308例，死亡2250例，死亡率高达66%。

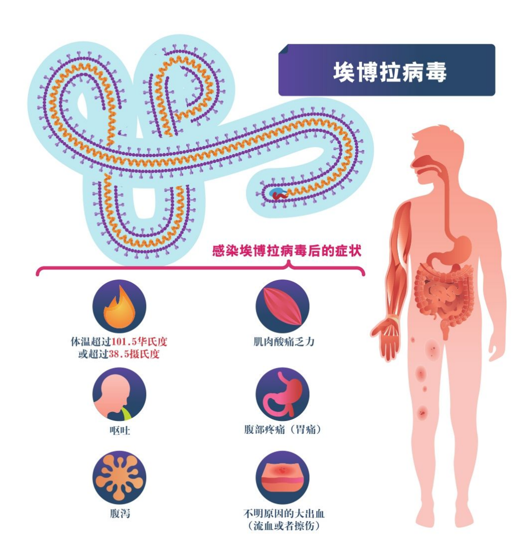
## 10.3　走进可怕的病毒

埃博拉病毒是一种不分节段的单股负链RNA病毒，由一个螺旋形状的核糖核壳复合体构成，它属于单股负链病毒目，丝状病毒科。丝状病毒科的病毒种类非常少，仅仅有3个属，分别是埃博拉病毒属、

马尔堡病毒属和2010年才命名的库瓦病毒属。

目前已经确定埃博拉病毒分为5种：扎伊尔型（Zaire ebolavirus,EBOV）、苏丹型（Sudan ebolavirus, SUDV）、莱斯顿型（Reston ebolavirus,RESTV）、塔伊森林型（Ta Forest ebolavirus, TAFV）和本迪布焦型（Bundibugyo ebolavirus, BDBV）。

埃博拉病毒主要通过病人的血液、唾液、汗水及分泌物、排泄物等传播，潜伏期为2～21天。绝大多数人会在5～10天发病。埃博拉病毒进入人体后，首先攻击的是单核细胞、巨噬细胞和树突状细胞，引起细胞变性坏死，凝结成块阻塞血管，导致组织液、血液大分子及细胞滞留。一旦病毒的膜和细胞膜发生了融合，该病毒就会进入细胞内部，释放出自己的RNA，随后在宿主体内疯狂繁殖，大量扩散，同时通过复制不断地攻击宿主体内的多个重要器官，导致器官发生内在损伤，同时产生内出血。当病毒侵入位于心脏、血管及肝肾等内脏器官的内皮细胞，引起血管和器官出现小的孔洞，血液成分顺着小孔流出。病变器官的坏死组织会从病人口中呕吐出来，病人死状非常恐怖，就仿佛身体内部的器官会逐渐“融化”。




感染埃博拉病毒后的症状

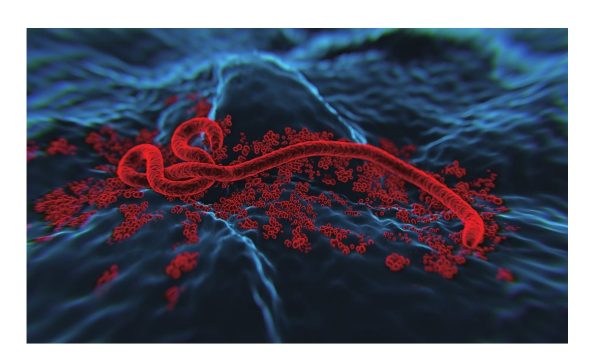
埃博拉病毒虽然致死率较高，但是它不能通过空气传播，前几次

造成的总死亡人数和天花、禽流感等相比较少，因此人们对它的重视程度不高，很多公司或者科研机构都觉得研发抗埃博拉的药物可能会入不敷出。

伴随着交通工具的发展，病毒可能在一天之内就从非洲传到美洲，说明这已经不单单是一个国家的事情，而是一个全球性的公共卫生事件。鉴于这种疾病的高死亡率，研制相应的疫苗已经迫在眉睫。

## 10.4　我们能获胜吗？

在人类与埃博拉病毒斗争的40年时间里，人类一直没有找寻到特效药，而以对症支持为主的治疗方案并不能有效降低病亡率。

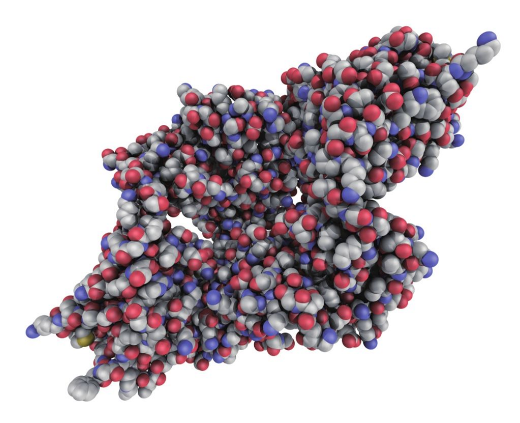


埃博拉病毒

在埃博拉病毒暴发伊始，治疗出血热尚无特效的方法，主要采取对症下药的辅助治疗，如维持水电解质平衡，预防和控制出血，控制继发性感染，维护各种脏器的功能，防止出现肾衰竭、出血等一系列并发症。尤其是恢复期患者的血清与免疫球蛋白可以作为疾病暴发阶段的经验性治疗药物。

2013—2016年埃博拉病毒的大暴发让疫苗研发正式提上日程。通过将灭活的埃博拉病毒注入实验动物体内，诱导机体产生免疫反应，但是这一思路在灵长类动物的实验中以失败告终。如果以减毒活的病毒直接进行人体实验，又怕对人体造成伤害，所以一直没有实施。

目前，抗埃博拉病毒的单抗疗法代表了抗击埃博拉出血热最有发展前景的方法之一，最具有代表性的人源化单克隆抗体制剂ZMapp已经开始用于治疗一些患者，但是这种方法生产的抗体不一定能够在紧急情况下快速响应。目前，科学界依然没有研制出广谱的抗病毒药物，埃博拉病毒感染只能依靠患者自身的免疫能力来进行抵抗，而相关治疗只能起到辅助的作用。

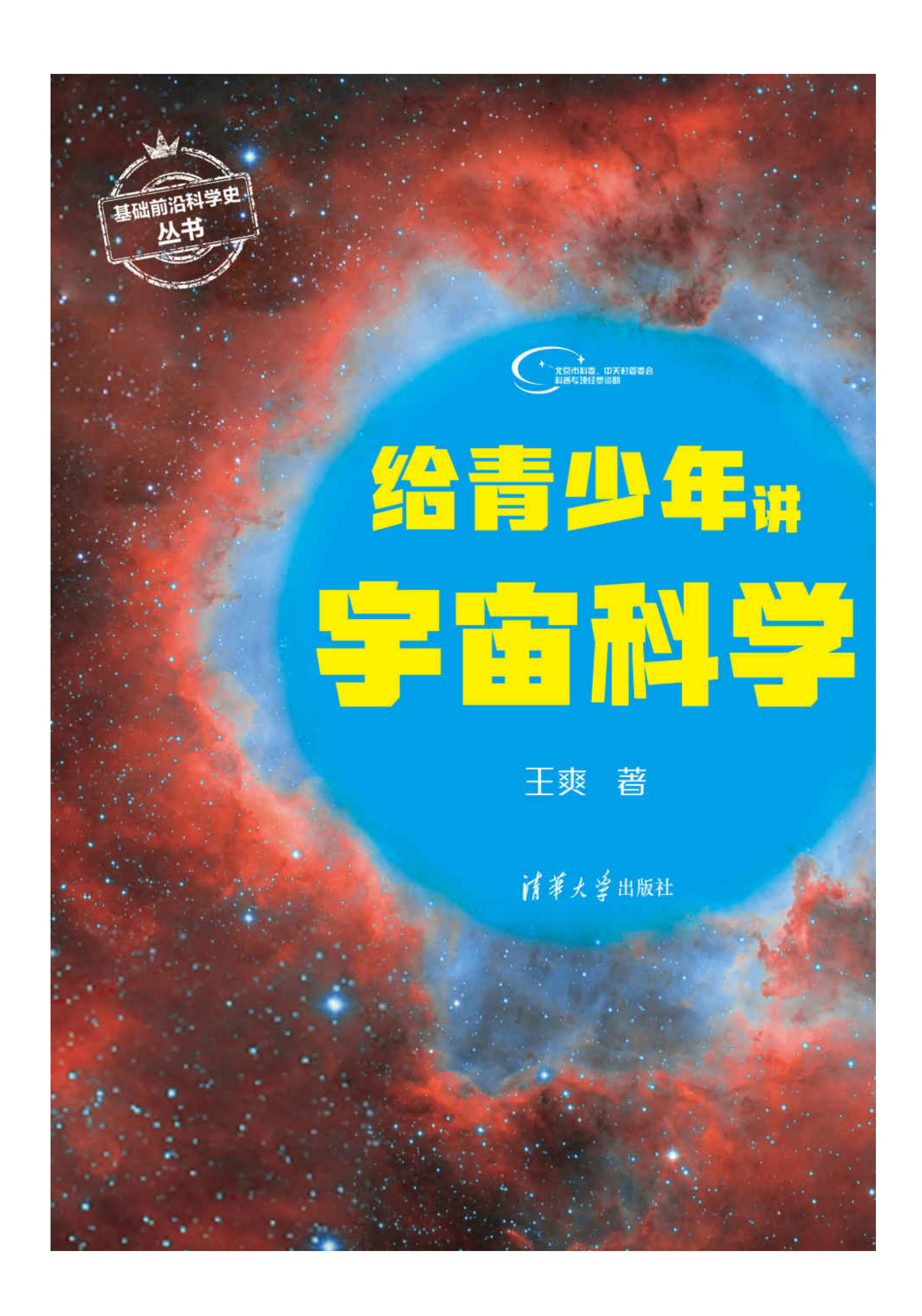


单克隆抗体

抗病毒药物仍在不断地研制之中，根据作用靶标的不同，可以分为抑制病毒入侵细胞和抑制病毒复制两类药物。2019年，科学家们最终发现了两种能够显著降低患病死亡率的药物：mAb114和REGN-EB3。这一研究成果有望改写埃博拉无药可治的历史，提升了人们对该传染病的防控能力和信心，因此该项成果被《科学》评为2019年十大科学突破之一。

2019年，美国食品药品管理局（Food and Drug Administration, FDA）批准上市了由默沙东公司研发的埃博拉减毒疫苗，该疫苗将埃博拉病毒的一段基因链接到水泡性口炎病毒基因上，在注射之后会让人体产生对埃博拉病毒的免疫能力，让我们初步看到了胜利的曙光。

与埃博拉病毒的斗争将会继续进行下去，相信人类必定会取得这场病毒战役的最终胜利！


基础前沿科学史
丛书
北京市科委、中关村管委会
科普专项经费资助
给青少年讲
宇宙科学
王爽 著
清華大學出版社

基础前沿科学史丛书

# 给青少年讲宇宙科学

王爽　著

清華大學出版社
北　京

图书在版编目（**CIP**）数据

给青少年讲宇宙科学 / 王爽著．—北京：清华大学出版社，2022.10

（基础前沿科学史丛书）

ISBN 978-7-302-61944-4

Ⅰ．①给…　Ⅱ．①王…　Ⅲ．①宇宙－青少年读物　Ⅳ．①P159-49

中国版本图书馆CIP数据核字（2022）第180869号

责任编辑：胡洪涛

封面设计：意匠文化·丁奔亮

责任校对：王淑云

责任印制：宋　林

出版发行：清华大学出版社

网　　址：http://www.tup.com.cn, http://www.wqbook.com

地　　址：北京清华大学学研大厦A座

邮　　编：100084

社总机：010-83470000

邮　　购：010-62786544

投稿与读者服务：010-62776969，c-service@tup.tsinghua.edu.cn

质量反馈：010-62772015，zhiliang@tup.tsinghua.edu.cn

印装者：三河市龙大印装有限公司

经　　销：全国新华书店

开　　本：165mm×235mm

印　　张：11

字　　数：119千字

版　　次：2022年11月第1版

印　　次：2022年11月第1次印刷

定　　价：55.00元

---

产品编号：097598-01

# 丛书序　给面向青少年的科普出版点一把新火

2022年是《中华人民共和国科普法》通过的第20年，在这样一个对科普工作意义不凡的年份，由北京市科学技术委员会（以下简称市科委）发起，清华大学出版社组织的“基础前沿科学史丛书”正式出版了。这套书给面向青少年的科普出版点了一把新火。

2022年9月4日，中共中央办公厅、国务院办公厅印发《关于新时代进一步加强科学技术普及工作的意见》，进一步强调“科学技术普及是国家和社会普及科学技术知识、弘扬科学精神、传播科学思想、倡导科学方法的活动，是实现创新发展的重要基础性工作”。科学技术普及是科技知识、科学精神、科学思想、科学方法的薪火相传——是“薪火”，也是“新火”。

市科委搭台，出版社唱戏，这套书给面向青少年的科普图书出版模式点了一把新火。市科委于2021年11月发布了“创作出版‘基础前沿科学史’系列精品科普图书”的招标公告，明确要求中标方在一年的时间内，以物质科学、生命科学、宇宙科学、脑科学、量子科学为主题，组织“基础前沿科学史”系列精品科普图书（共5册）出版工作；同步设计制作科普电子书；通过网络媒体对图书进行宣传推广等服务内容。这些服务内容以融合出版为基础，以社会效益为初心。服务内容的短短几句话，每一句背后都是特别繁复的工作内容。想在一年的时间内，尤其是在2022年新冠肺炎疫情期间，完成这些工作的难度可想而知，然而秉承“自强不息，厚德载物”的清华大学出版社的出版团队做到了。

中国科学家，讲好中国故事，这套书给面向青少年的科普图书选题内容点了一把新火。中国特色社会主义进入新时代，新一轮科技革命和产业变革正在深入发展，基础前沿科学改变着人们的生产生活方式及思维模式。《中华人民共和国国民经济和社会发展第十四个五年规划和2035年远景目标纲要》提出：在事关国家安全和发展全局的基

础核心领域，制定实施战略性科学计划和科学工程。物质科学、生命科学、宇宙科学、脑科学、量子科学等领域，迫切需要更多人才参与研究，而前沿科学人才的建设培养，要从青少年抓起。这5本书的作者都是中国本土从事相关专业领域工作的科学家，这5本书都是他们依托自己工作进行的原创性工作。虽然内容必然涉及科学史的内容，但中国科学家尤其是近些年的贡献也得到了充分展示。

初心教育，润物无声，这套书给面向青少年的科普图书科普创作点了一把新火。习近平总书记提出：科技创新、科学普及是实现创新发展的两翼，要把科学普及放在与科技创新同等重要的位置。因此，针对前沿科技领域知识的科普成为重点。如何创作广受青少年欢迎的优秀科普图书，充分发挥科普图书的媒介作用，帮助青少年树立投身前沿科学领域的梦想，是当前科普出版工作的重点之一，这对具体的科普创作方法提出了要求。这套书，看得出来在创作之初即统一了整体创作思路，在作者进行具体创作时又保持了自己的语言习惯和科普风格。这套书充分体现了，面向青少年的科普图书创作，应该循序渐进，张弛有度，绘声绘色，娓娓道来，以科学家的故事吸引他们，温故科学家的研究之路，知新科学家的科研理念，以科学精神润物细无声。

靡不有初，鲜克有终。2022年10月16日，习近平总书记在中国共产党第二十次全国代表大会报告中强调“教育、科技、人才是全面建设社会主义现代化国家的基础性、战略性支撑”。且将新火试新茶，诗酒趁年华。希望清华大学出版社的这套“基础前沿科学史丛书”为广大青少年推开科学技术事业的一扇门，帮助他们系好投身科学技术事业的第一粒扣子，在全面建设社会主义现代化强国的新征程上行稳致远。

中国工程院院士

清华大学教授

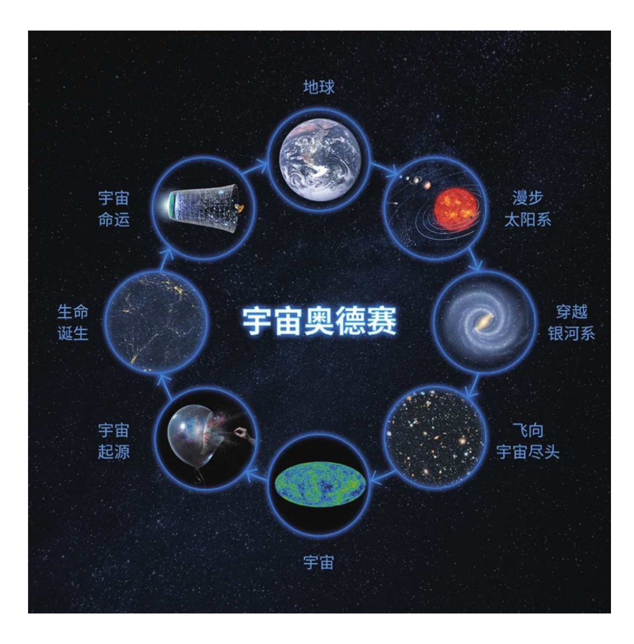


# 前言

2017年夏天，我开始规划一场名为“宇宙奥德赛”的环游宇宙之旅。

下图就是宇宙奥德赛之旅的规划图。旅程的前半段是空间之旅。我们从地球出发，按照由近及远的顺序，依次游历以太阳系为代表的行星世界、以银河系为代表的恒星世界和银河系之外的星系世界，最终到达宇宙的尽头，同时也是时间的起点。旅程的后半段则是时间之旅。我们会从宇宙创生的时刻出发，在时间长河中顺流而下，依次探寻宇宙起源、生命诞生和宇宙命运的奥秘，并最终回到今天的地球。旅程结束后，我们就能真正了解人类最终极的三大哲学问题（我是谁？我从哪里来？我将往何处去？）的答案。

图中除“地球”和“宇宙”以外的每一个圆，都代表一本科普书。换言之，我计划用6本科普书的篇幅，来完成这场宇宙奥德赛之旅。




“宇宙奥德赛”系列的前两本书，《宇宙奥德赛：漫步太阳系》和

《宇宙奥德赛：穿越银河系》，目前的豆瓣评分分别为9.1和9.2，双双跻身于读者口碑最好的中国本土科普书之列。此系列的第三本书，《宇宙奥德赛》也将于今年年底出版。

今年年初，清华大学出版社的胡洪涛主任问我，能不能写一本为青少年介绍宇宙科学的科普书。很自然地，我就想到可以写一个宇宙奥德赛之旅的简略版，也就是本书。

本书的逻辑主线，就是上图所展示的宇宙奥德赛之旅。我从这趟环游宇宙之旅中精选了12个最重要的主题，包括日心说和地心说、天文距离测量简史、标准烛光、银河系的大小、可观测宇宙的大小、宇宙膨胀、暴涨、宇宙大爆炸、宇宙微波背景、恒星的一生、暗物质与暗能量、宇宙的终极命运。前6个主题，描述了前半段的宇宙空间之旅；而后6个主题，则描述了后半段的宇宙时间之旅。对这12个精选主题的阅读，可以为你构建一个关于宇宙的知识体系的骨架。

写作手法上，本书有两个最核心的特点：1．内容可视化。全书几乎没有数学公式，所有科学知识都转化成了可视化的物理图象，再用通俗易懂的比喻来加以解释。2．故事驱动。为了增加趣味性，书中穿插了大量的科学家的逸闻趣事。此外，我也借鉴了评书的创作技巧，在每一章的结尾都留下了一个承前启后的科学问题。相信你能感受到此书中倾注的心血和诚意。

准备好了吗？那我们就开始这场环游宇宙之旅吧。

# 目 录

# 1　日心说和地心说

估计很多人都听过这样的说法：1543年，波兰大天文学家哥白尼（Kopernik）提出了日心说，从而一举打破了长期居于绝对统治地位的地心说，实现了天文学的伟大变革。

我要告诉你的是，这种说法是错的。地心说和日心说的斗争历史，其实异常的曲折和漫长。

所以这门宇宙科学的第一节课，我就来讲讲地心说和日心说斗争的曲折历史。

早在2000多年前，古希腊人就提出了地心说和日心说。

人类很早以前就观察到，日月星辰似乎都在周而复始地绕着地球旋转。所以绝大多数人都认为，地球位于整个宇宙的中心。地心说就是这么起源的。

图1就展示了地心说的基本图像。地球位于宇宙的中心，就像是位于一个城市的市中心。地球周围有7个圆形的轨道，相当于城市的7环：按从内到外的顺序，依次是月球、金星、水星、太阳、火星、木星和土星。这7个天体都在各自的圆形轨道上，沿着相同的逆时针方向绕地球旋转。在7环之外，还有一个大天球，其他的星星就散布在这个天球之上。

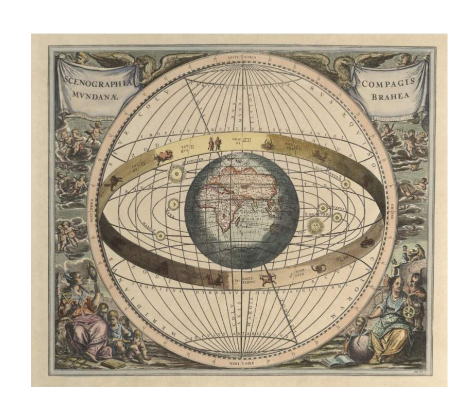
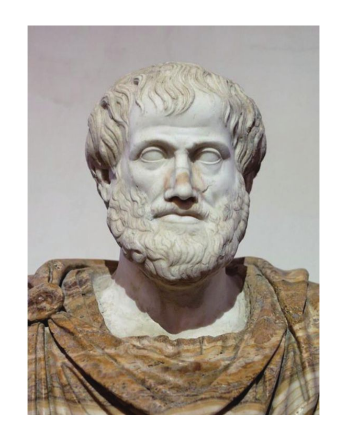


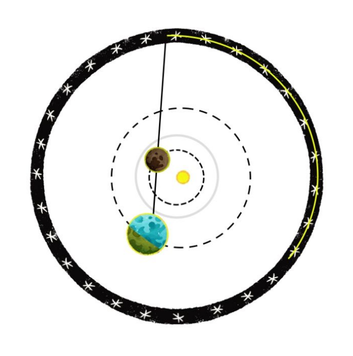
图1

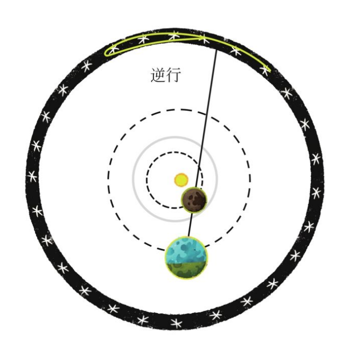


顺便说一下。地心说最大牌的支持者，是古希腊大哲学家亚里士多德（Aristotle）。他提出了下面这个思想实验，来论证地球必须静止不动：一个人原地向上跳，如果地球在运动，那么当此人落地时，就无法落回原地，而会落到其他的位置。但真实情况是，此人肯定会落回原地。所以亚里士多德就宣称，地球一定是静止不动的。

当然，以我们今天的眼光来看，亚里士多德的论证无疑是错的。聪明的读者，你猜到他到底错在哪里了吗？

不过，尽管有亚里士多德这样的超大牌支持者，地心说在古希腊时代并没有一统天下。因为它有一个很致命的缺陷，那就是行星逆行。之前说过，地心说认为，所有的行星都必须沿逆时针方向绕地球旋转。但是人们很早就发现，很多行星的旋转方向经常会突然变成顺时针。这种行星旋转方向突然改变的现象，就是所谓的行星逆行。这对早期的地心说来说，可谓是致命一击。






所以，就有了一群反对地心说的人。其中的代表人物，是古希腊天文学家阿利斯塔克（Aristarchus）。



阿利斯塔克的理论源于一个哲学命题。他认为，世界的本源是火。既然火是万物的本源，那么火一定得位于全宇宙最重要的位置，也就是中心。所以，宇宙的中心一定是太阳。这就是日心说的起源。

图2就展示了日心说的基本图像。太阳位于宇宙的中心。在太阳的周围还有6个圆形轨道，相当于城市的6环：按从内到外的顺序，依次是水星、金星、地球、火星、木星和土星。这6颗行星都在各自的圆形轨道上，沿着相同的逆时针方向绕太阳旋转。而在6环之外，则是一个散布着各种星星的大天球。

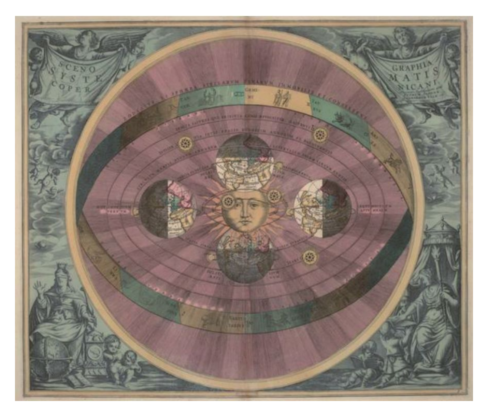


图2

与地心说相比，日心说最大的优势是它可以很轻松地解释行星逆行的现象。所以尽管阿利斯塔克远远不如亚里士多德大牌，日心说依然与地心说分庭抗礼了将近500年。

直到公元140年，一个超级天才的横空出世，才彻底打破了地心说和日心说之间的平衡。此人就是古罗马帝国大天文学家克罗狄斯·托勒密（Claudius Ptolemaeus）。



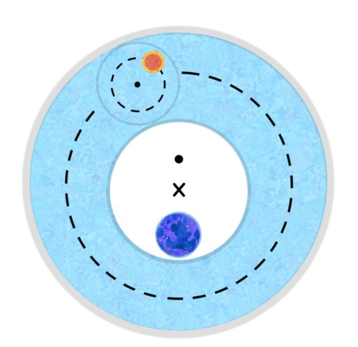
托勒密为什么能打破地心说和日心说之间的平衡呢？答案是，他对最早期的地心说进行了修改，从而破解了地心说无法解释行星逆行现象的世纪难题。

为了介绍托勒密的理论，我要拿一个在现实世界里很常见的事物进行类比，那就是游乐园里的旋转咖啡杯。

一般而言，旋转咖啡杯的中心会有一个茶壶。在茶壶的周围会有一个大圆的轨道，上面分布着一些圆形的咖啡杯。除了能绕茶壶旋转以外，咖啡杯自己也可以旋转。当机器开动以后，游客会坐在咖啡杯的边缘，既绕着茶壶的大圆轨道旋转，又绕着咖啡杯的小圆轨道旋转。

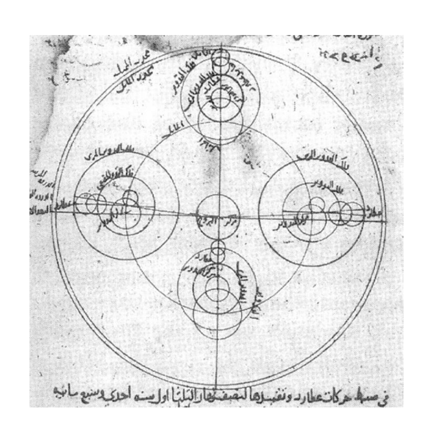


托勒密的解决之道和这个旋转咖啡杯的图像非常类似。他认为，地球并不在宇宙的正中心，而是与真正的中心有一个很微小的偏离。更关键的是，包括金星、水星、火星、木星、土星在内的五颗行星，都像是乘坐旋转咖啡杯的游客：首先，行星会在一个叫本轮的小圆上旋转，就像是咖啡杯的小圆轨道；然后，本轮圆心又会在一个叫均轮的大圆上旋转，就像是茶壶的大圆轨道。因此，行星的运动轨迹是由本轮和均轮这两个圆周运动组合而成的。

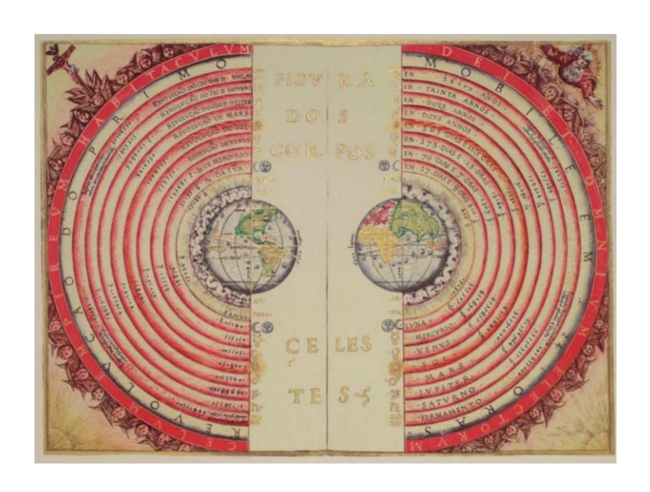


引入这个类似于旋转咖啡杯的本轮-均轮体系，能让行星的运动轨迹变复杂。此外，这个本轮-均轮体系还可以不断地拓展。例如，你可以把原来的本轮视为一个新的均轮（相当于把原来的咖啡杯当成新的

茶壶，即第2层均轮），然后在它周围画更小的本轮（即第2层本轮）。也可以把第2层本轮当成第3层均轮，然后在它周围画第3层本轮。随着层数的不断增加，行星的运动轨迹就会变得越来越复杂。这样一来，行星逆行的现象就很容易解释了。



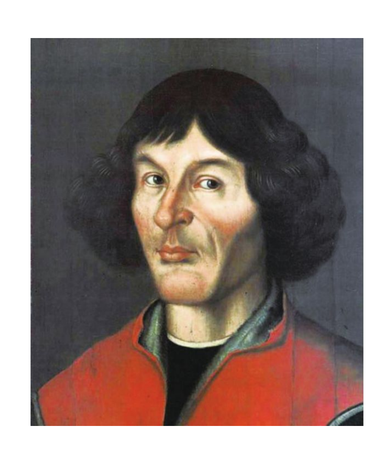
简单地总结一下。通过引入类似于旋转咖啡杯的本轮-均轮体系，托勒密在地心说的理论框架下，成功地解决了行星逆行的世纪难题。这样一来，经过托勒密改良的地心说（地球很靠近宇宙的中心，月球和太阳还是以圆形轨道绕地球旋转，五颗行星则位于多层嵌套的旋转咖啡杯上），就一举击败了与自己缠斗数百年的日心说。




而到了13世纪，一个叫托马斯·阿奎纳（Thomas Aquinas）的神学家让地心说的地位更上一层楼。他把地心说融入天主教神学体系

中。一旦质疑地心说，就相当于质疑天主教本身。这样一来，地心说就一统天下了。

但300年后，另一个人的出现，让胜负的天平再次发生了偏移。此人就是波兰大天文学家哥白尼。1543年，他出版了一本宣扬日心说的书，叫《天体运行论》。正是这本书，让被遗忘了1400多年的日心说死灰复燃。



有趣的是，这本书其实差点儿就被哥白尼带进坟墓。

这是因为哥白尼的本职工作是一名天主教教士，所以他心里很清楚，公开宣扬日心说会得罪整个天主教会。因此，尽管他在40岁的时候就已经开始在一个小圈子里宣扬自己的理论，却始终不肯著书出版。

那这本书后来为何又出版了呢？是由于一个不速之客。此人就是奥地利数学家格奥尔格·雷蒂库斯（Georg Rheticus）。

1539年，雷蒂库斯了解到了哥白尼改良后的日心说，顿时觉得醍醐灌顶。为此，他专门跑到波兰去找哥白尼，想游说哥白尼著书出版此理论，结果却吃了闭门羹。

但雷蒂库斯锲而不舍。此后两年多的时间，他就像块牛皮糖，死死粘住了哥白尼，反复游说哥白尼一定要著书立说。最后，哥白尼终于招架不住，交出了《天体运行论》的书稿。

拿到书稿以后，雷蒂库斯就开始寻找愿意资助出版此书的出版

商。一年后，他找到了一个愿意出钱的出版商。一切终于走上了正轨。

由于在《天体运行论》中有大量的公式和图表，必须得有一个专家来做此书的编辑，以保证内容的准确性。雷蒂库斯做了半年的编辑工作。但他后来有急事，不得不中途离开。临走前，雷蒂库斯找了个继任者，叫奥西安德尔（Osiander）。

既荒诞又搞笑的事情来了。

奥西安德尔接手了编辑工作以后才发现，这本书竟敢公然反对地心说，顿时觉得自己上了一条贼船。为了避免池鱼之殃，他干了一件今天的编辑连想都不敢想的事情：瞒着哥白尼和雷蒂库斯，伪造了一篇前言，宣称此书“并不是一种科学的事实，而是一种富于戏剧性的幻想”。

不过这个伪造前言的恶行并没有受到追究。因为当此书正式出版的时候，哥白尼已经去世了。

很多中小学科学读物讲到这段历史的时候，都会说：“哥白尼提出的日心说一举打破了地心说的统治地位，实现了天文学的伟大变革。”

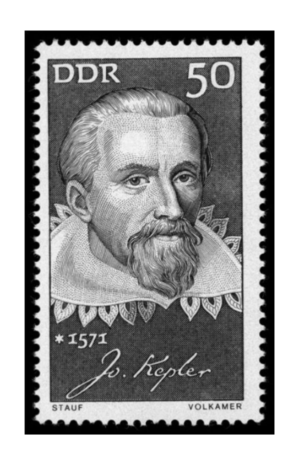
这个说法是错的。

在哥白尼重新提出日心说后的大半个世纪里，地心说的地位依然坚如磐石。直到17世纪初，两个科学巨人的横空出世，才让胜利的天平倒向日心说。

先讲讲日心说为何迟迟得不到学术界的认可。原因在于，它无法解释火星的轨道异常。

按照日心说的观点，太阳位于宇宙的中心；其他的行星，都沿着圆形轨道绕太阳公转。但人们后来发现，火星的运动轨道相当诡异，并不是一个完美的圆。哥白尼本人也意识到了这个问题。无奈之下，他引入了托勒密的本轮-均轮体系，把火星也放在一个“旋转咖啡杯”上。但这样一来，日心说就失去了它相对于地心说的最大优势：数学上简单明了。

破解这个难题的是我们要讲的第一个科学巨人——德国大天文学家约翰内斯·开普勒（Johannes Kepler）。



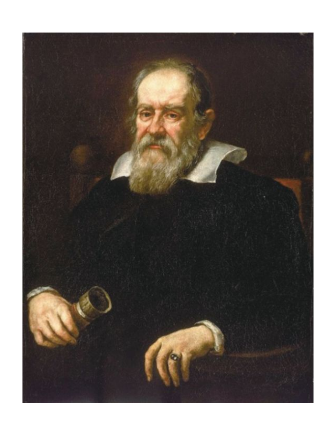

开普勒被后人称为“天上的立法者”。他之所以有这样的盛名，是因为他在17世纪初提出了著名的“行星运动三定律”。在这三条定律中，最具颠覆性的是第一定律。它说的是，行星绕太阳公转的轨道并不是圆，而是椭圆。这就解释了火星的轨道异常。因为火星绕太阳旋转的轨道，正是椭圆。

开普勒的发现让日心说有了和地心说平起平坐的实力。但要想真正打败地心说，必须发现一种特殊的自然现象；这种自然现象用地心说根本说不通，用日心说却能完美地解释。

发现这种自然现象的人，就是我们要讲的第二个科学巨人，此人就是被后人称为“现代科学之父”的伽利略（Galileo）。

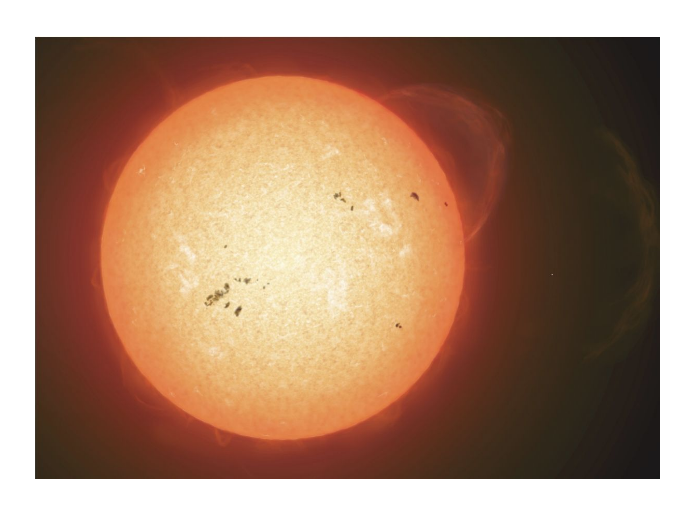
伽利略发现这种自然现象的故事，得从一个不相干的人讲起。

1608年，一个荷兰眼镜店老板偶然发现用两块前后放置的镜片可以看清远处的物体，进而造出了人类历史上的第一架望远镜。这个消息传到了意大利，立刻引起了伽利略的浓厚兴趣。



1609年，伽利略制造了一个质量更好的望远镜，能把远处的物体放大30多倍。然后，他做了一件意义非凡的事情：把这个望远镜指向了太空。

这个举动，宣告了现代天文学的诞生。

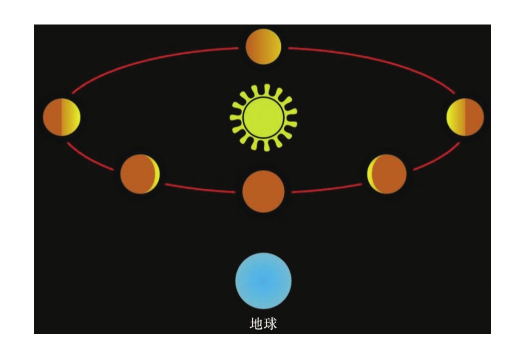


伽利略第一次用望远镜仰望太空的心情，应该和阿里巴巴第一次看见满山洞财宝的心情差不多。用这个望远镜，他发现了很多人类前所未见的景象，比如太阳黑子、月球环形山和木星卫星。其中最有影响力的发现，直接导致了地心说的衰落和日心说的崛起，那就是金星盈亏。

什么是金星盈亏呢？我们不妨用月球盈亏来做一下类比。

相信很多人都知道，月球是有盈亏的。为什么月球会有盈亏呢？因为月球本身不发光，只能反射太阳光。由于月球一直在绕地球旋转，它既能跑到地球和太阳之间，也能跑到地球的背后。如果月球跑到了地球和太阳之间，它就会把后面射来的太阳光挡住，让我们无法看到它，这就是月球的“亏”；如果月球跑到了地球的背后，它就可以完全地反射太阳光，让我们看到一轮圆月，这就是月球的“盈”。

与月球不同的是，金星无法跑到地球的背后。不过，它有可能跑到太阳的背后。如果金星跑到了地球和太阳之间，它就会挡住后面射来的太阳光，让我们看不到它，这就是金星的“亏”；如果它跑到太阳的背后，就可以完全地反射太阳光，让我们看到一个最圆最亮的金星，这就是金星的“盈”。




知道了金星盈亏的概念，我们就可以来讲讲如何判断地心说和日心说的对错了。问题的关键在于，金星到底是绕地球旋转还是绕太阳旋转。

天文观测表明，金星一直都在太阳周围活动。在地心说中，金星一直在绕地球旋转；要想解释金星总在太阳周围活动的观测结果，金星和太阳就必须以差不多的角速度绕地球旋转。在这种情况下，金星就只能一直处于地球和太阳中间，永远不可能出现“盈”的状态。

而在日心说中，金星一直绕太阳旋转，所以能自然而然地解释为什么金星总在太阳周围。更重要的是，在这种情况下，金星也可以跑到太阳的背后，从而出现“盈”的状态。

所以地心说和日心说就有一个最本质的区别。在地心说中，金星

绕地球旋转，因此只能“亏”不能“盈”；而在日心说中，金星绕太阳旋转，因此既能“亏”又能“盈”。这样一来，通过观察金星能否出现“盈”的状态，就可以判断它到底绕着谁旋转，进而判断地心说和日心说的对错。

1610年，伽利略用他自制的望远镜，真真切切地看到金星确实出现了“盈”的状态。在一封寄给朋友的信中，伽利略富有诗意地写道：“爱之母（金星）正在效仿辛西娅（月亮女神）的风姿。”同一年，他把这个发现写进了自己的传世名著《星际信使》，从而敲响了地心说的丧钟。

打败地心说后，日心说就成了天文学界的经典理论；它的统治，一直延续到了20世纪初。

以今天的眼光来看，日心说的错误也是很明显的：宇宙的中心，当然不可能是太阳。那么，为什么今天看来错误相当明显的日心说，却能继续统治天文学界长达300年？

欲知详情，请听下回分解。

# 2　天文测距简史

上节课的结尾，我们提出了这样一个问题：为什么今天看来错误明显的日心说，却能统治天文学界长达300年？要想回答这个问题，我必须先讲讲人类在20世纪以前测量遥远天体距离的历史。

图3就展示了20世纪前的天文距离测量的核心历史。简单地说，人类实现了“三级跳”：其中的一级跳，是利用几何学的知识，测出了地球的直径；而二级跳，是以地球直径为尺，通过观察金星凌日现象，测出了太阳和地球的距离（即日地距离）；至于三级跳，是以日地距离为尺，基于三角视差方法，测出了更遥远的恒星的距离。

听起来是不是有点儿云里雾里？那我接下来就详细地介绍一下。

先说一级跳：如何测出地球的直径。

要测地球的直径，就相当于要测地球的周长。那么，地球的周长该怎么测量呢？



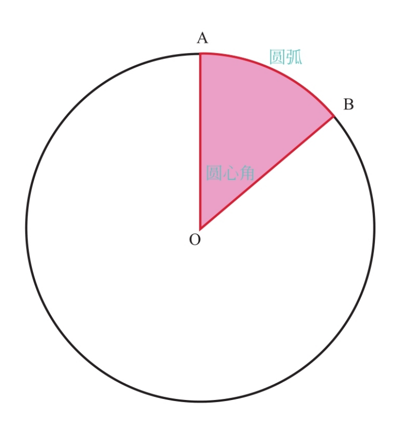

图3

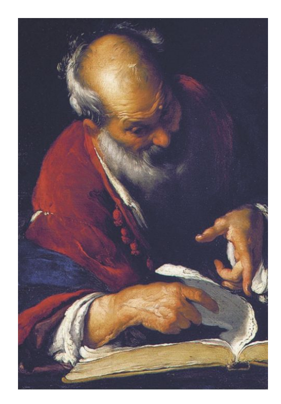
按常理来说，这根本就是个不可能完成的任务。因为地球表面70%的区域都被海水覆盖。由于海洋的阻隔，根本不可能沿地球赤道走一圈，进而用尺子量出地球的周长。

不过，有人想出了一个另辟蹊径的妙招：先用尺子量出地球表面一段圆弧的长度，再想办法确定这段圆弧对应的圆心角（如下图所示）。这样一来，就能确定这段圆弧相对于地球周长的比例，进而算出地球的周长和直径。

世界上第一个准确测出地球周长和直径的人，是古希腊大哲学家埃拉托色尼（Eratosthenes）。

埃拉托色尼被后世称为“地理学之父”。像经度和纬度的概念，都是他最早提出的。



但即使是这样的顶级“大牛”，早年过得也不是很如意。在前半生，他一直被别人叫作“千年老二”。这是因为有个人样样都比他厉害。此人就是宣称“只要给我一个支点，我就能撬起地球”的阿基米德(Archimedes)。

眼看自己没有斗过阿基米德的希望，埃拉托色尼选择了远走他乡。他接受了埃及国王托勒密一世的邀请，跑到埃及做了亚历山大图书馆的馆长。

托勒密一世希望，能在有生之年看到亚历山大图书馆变成全世界最大的图书馆。这并不是一件简单的事。因为当时的亚历山大图书馆与雅典的图书馆还有很大的差距。

埃拉托色尼用一种很奸诈的手段，完成了托勒密一世的夙愿。他先向雅典的图书馆付了一大笔钱，把它的大量藏书都借到埃及展览。然后，他又找了一大批馆员来临摹藏书副本。这些副本临摹得特别好，几乎达到了以假乱真的程度。所以，最后还书的时候，埃拉托色尼就只还了这些藏书的副本，而把真品留在了自己的图书馆里。靠着这样的手段，亚历山大图书馆很快成了当时全世界最大的图书馆。

在当馆长之余，埃拉托色尼也会利用图书馆的资源进行学术研究。他一生中最有名的研究工作，就是对地球周长的测量。他是怎么测量的呢？答案是，用到了一口特殊的井。

埃及南部有一个叫赛伊尼的城市。这个城市有一口有名的深井：在夏至日的正午时分，太阳光恰好能直射到这口深井的井底（之所以会有这样的现象，是因为这口井恰好位于北回归线上）。这个现象很有名，每年夏至日都能吸引到不少的游客。埃拉托色尼发现，它还能用来测量地球的周长。

听起来好像有点儿不知所云？其实只要用一点简单的几何学知识，就可以把它说清楚。

图4就是埃拉托色尼测量地球周长的原理图。此图展示了夏至日的正午时分，太阳光照射埃及的情况。

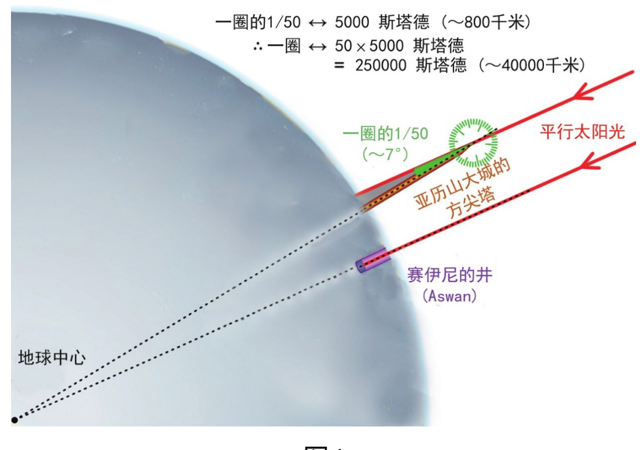



图4

图4中的紫色圆柱就是赛伊尼的那口深井。前面说过，在夏至日的正午时分，红色平行线所代表的太阳光可以直射到这口深井的井底。这意味着，这束直射井底的太阳光恰好可以穿过地球的球心。在同一时刻，埃拉托色尼在亚历山大城测量一个很高的方尖塔（即橙色长条）的阴影长度，并以此算出这个方尖塔与太阳光之间的夹角（即绿色夹角）约为7.2°。利用初中几何的知识，就可以知道这个夹角恰好等于赛伊尼和亚历山大城之间的那段圆弧相对于地球球心的圆心角。因为环绕地球一圈的圆弧角度是360°，所以这两座城市之间的距离约为地球周长的1/50。

知道了赛伊尼和亚历山大城之间的圆弧占整个地球周长的比例后，接下来就只需量出这段圆弧的长度（即赛伊尼的井和亚历山大城方尖塔之间的距离）。为此埃拉托色尼专门派出了一支商队，用尺子一点点地量出两地之间的距离约为5000斯塔德（1斯塔德＝157米）。由此可以算出，地球的周长约为250000斯塔德。

古埃及人的1斯塔德，相当于现代人的157米。所以埃拉托色尼的测量结果，相当于今天的39250千米。拿它和今天的结果对比一下。根据地球卫星的测量结果，地球的周长是40076千米。换言之，埃拉托色尼在2200多年前测出的地球周长与今天最精确的测量结果，只有区区2%的误差。

一旦知道地球周长，就可以算出地球直径，大概是12724千米。这样一来，人类就完成了天文距离测量的一级跳。

再说二级跳：如何测出日地距离。

很明显，日地距离就更不可能用尺子直接测量了。所以，同样得另辟蹊径，用几何学的办法，算出太阳和地球之间的距离。

最早想到这个办法的人，是英国著名天文学家埃德蒙多·哈雷（Edmond Halley）。

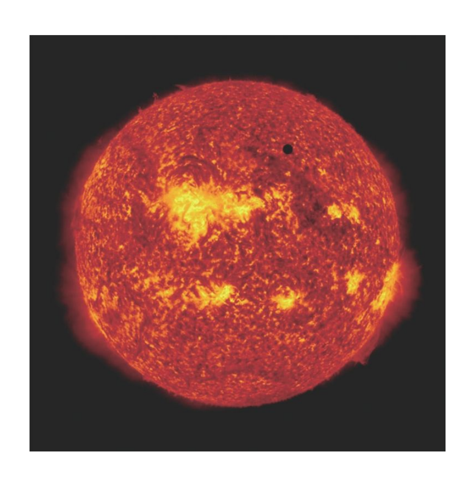


哈雷是一个典型的少年天才。19岁那年，还在读本科的哈雷就成了英国首任皇家天文学家约翰·弗兰斯蒂德(John Flamsteed)的助手。在弗兰斯蒂德的资助下，哈雷跑到南大西洋的一个小岛上，建立了南半球的第一个天文台。随后，他在那里绘制了全世界第一张南天星表，并因此当选为英国皇家学院院士。那一年，哈雷只有22岁。

哈雷后来又做出了一大堆杰出的贡献。例如，他算出了哈雷彗星的轨道，并预言它每隔76年就会回归一次；他也制作了世界上第一张气象图，发明了世界上第一个潜水钟，还写了世界上第一篇关于人寿保险的论文。

而对天文学影响最深远的，是他在1716年发表的一篇论文。在这篇论文中哈雷指出，通过对金星凌日的仔细观察，就可以测出地球与太阳间的距离。

什么是金星凌日呢？为了便于理解，我还是用月球来进行类比。我们知道，月球有时能跑到地球和太阳的中间，挡住太阳光射向地球的路线，这就是日食。同样的道理，金星有时也能跑到地球和太阳的中间。这时在地球上观测，就可以看到一个小黑点在太阳表面缓慢地穿行。一般来说，这个小黑点要花几小时才能通过太阳的表面。这个现象就是“金星凌日”。



知道了什么是金星凌日，我们就可以看图说话，来讲讲用金星凌日测量日地距离的原理图了。

图5中左边的黄色大球代表太阳，右边的蓝色小球代表地球，太阳与地球之间的黑色圆点代表金星。

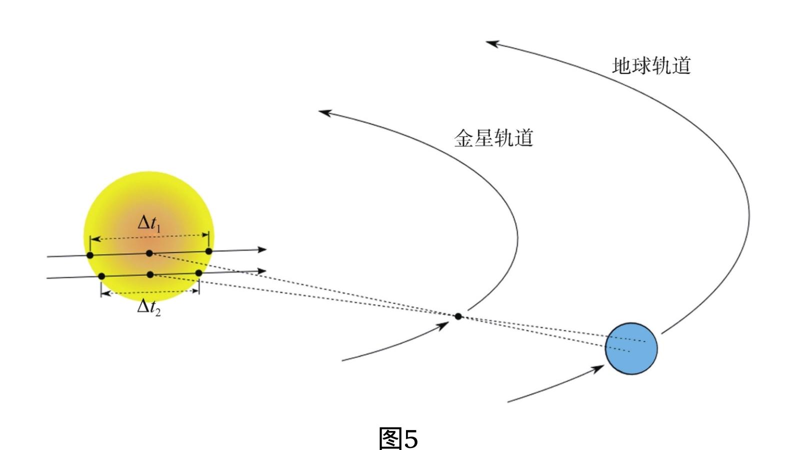



图5

哈雷指出，可以在地球上两个经度相同而纬度不同（最好纬度相差较大）的地点同时观测金星凌日。从图5中很容易看到，在这两地看到的金星凌日的轨迹（即太阳上的两条黑色实线）截然不同，从而在这两地测出的金星凌日持续时间会出现一定的差异。通过比较两者之间的时间差，能算出金星与地球上这两地所构成的等腰三角形的顶角。知道了这个顶角，再利用地球直径和两地纬度算出这两地之间的距离，就可以知道地球与金星之间的距离。

知道了地球与金星之间的距离，再利用开普勒行星运动第三定律（即行星运动周期的平方与其椭圆轨道半长轴的立方成正比），就能算出日地距离了。

哈雷以一己之力，提出了利用几何学和金星凌日测量日地距离的方法。那接下来，就该好好观测金星凌日了。

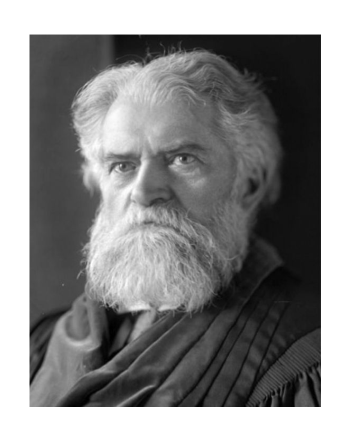


但问题在于，金星凌日是一种极为罕见的天文现象。它总是成对出现，且两次金星凌日之间总是相隔8年。如果这两次金星凌日都没赶上，那就只能再等100多年，才能看到下一对的金星凌日。

1761年，也就是哈雷去世的19年后，天文学家终于等到了观测金星凌日的机会。但遗憾的是，由于摄像技术的局限性，没有人能够测出日地距离的准确值。

又过了100多年，终于有人完成了精确测量日地距离的壮举。此人就是美国天文学家西蒙·纽康（Simon Newcomb）。

1882年，纽康做了一件相当大手笔的事情。他组织了整整8支科考队，分赴世界各地，来观测当年发生的金星凌日。通过整合8支科考队的数据，他测出日地距离应为1.4959亿千米。这个100多年前测出的数值，与今天的测量结果相差无几。



现在天文学界普遍接受日地距离约为1.5亿千米。这个距离，也被称为一个天文单位。

由此，人类完成了天文距离测量的二级跳。

最后说说三级跳：如何基于三角视差法，测量更遥远天体的距离。

什么是视差呢？为了更好地理解这个概念，我们不妨做个小实验。先伸出一根手指，放在靠近鼻子的地方；然后轮流闭上左右眼，每次都只用一只眼睛来看手指。你会发现，手指相对于背景的位置发生了很明显的偏移。这种由于观察者位置改变而导致被观察物体位置发生偏移的现象，就是视差。

现在把手指放在比较远的地方，重复这个实验。你会发现，将手指放远以后，它相对于背景的位置偏移会变小。反过来说，被观察物体的视差越小，它离我们的距离就越远。

顺便多说一句。电影院里放的3D电影之所以能呈现出立体感，就

是利用了视差的原理。

知道了什么是视差，我们就可以讲讲如何用三角视差法测量遥远天体的距离了。图6就是三角视差法的原理图。

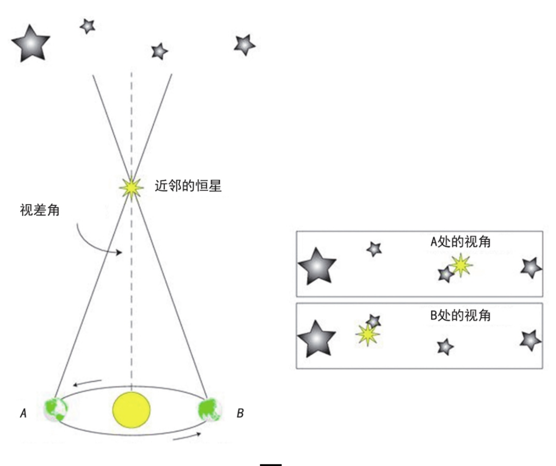



图6

我们知道，地球每年会绕太阳转一圈。如果地球在某个时刻运动到图中的$A$点，那么半年之后它就会到达离$A$点最远的$B$点。现在把$A$点和$B$点当成是一个人的左眼和右眼，然后分别在这两个地方观察一颗远处的星星。很明显，由于视差的缘故，这颗星星在遥远天幕上的位置会发生改变。利用这个位置的改变，能算出此星星与$A$、$B$两点所构成的等腰三角形的顶角，也就是所谓的周年视差角。这样一来，只要知道了日地距离，就能知道$A$、$B$两点的间距；而用$A$、$B$两点的间距除以该星星的周年视差角，就可以算出我们到这颗星星的距离。

这种以日地距离为尺，并用几何学知识测量遥远天体距离的方法，就是三角视差法。在20世纪以前，这是人类所知最强大的测量遥远天体距离的方法。



但要命的是，三角视差法依然是一种能力非常有限的测距方法。

举个例子，2006年，4位天文学家在《科学》杂志上发表了一篇论文。他们用三角视差法测量了地球与银河系英仙臂中的一团分子云的距离，结果是6370光年（1光年 $=9.46\times10^{12}$ 千米，也就是说，要想从地球去那里，就连光也要走6370年）。这个发现，创造了当时用三角视差法测到的最远距离的记录。换句话说，这几乎就是三角视差法的测距能力极限。

我们现在已经知道，在20世纪以前，为了测量遥远天体的距离，天文学界进行了三级跳，最远可以跳到6000多光年。但问题是，我们生活的这个银河系，其直径至少有10万光年！

现在终于可以回答本节课开头提出的那个问题了：为什么今天看来错误明显的日心说，却能统治天文学界长达300年？答案是，就连人类当时所知的最强大的天文测距方法，也完全不具备测量整个银河系的能力。所以，人类就陷入了“不识庐山真面目，只缘身在此山

中”的困境。

正因为如此，一直到20世纪初，人类都普遍相信：银河系就是宇宙的全部，而太阳就位于宇宙的中心。

要想突破这个困境，进而打破哥白尼日心说的禁锢，就必须找到一种全新的天文测距方法，其能够测量非常遥远的天体的距离。

是谁找到了全新的天文测距方法？又是谁敲响了哥白尼日心说的丧钟？

欲知详情，请听下回分解。



# 3　标准烛光

上节课的结尾，我们提出了这样一个问题：是谁敲响了哥白尼日心说的丧钟？

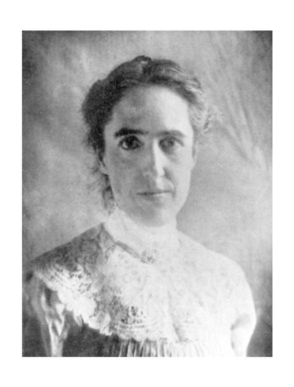
在回答这个问题之前，我想先讲一个故事。

1920年，美国政府在马萨诸塞州的剑桥郡搞了一次人口普查。有个人口普查员负责剑桥郡一个比较贫穷的社区。在那里，他遇到了一对相依为命的母女。

比较特殊的是那个女儿，因为她是一个聋女，费了好大劲才弄清人口普查员的来意。不过，她后来一直都很配合。

最后，她被问到自己目前从事什么职业。她的回答是“科学家”。

那个人口普查员当时就笑出声了。在那个年代的美国，科学家完全是男人的专属领地，几乎没有女性能够拿到博士学位。所以他根本无法想象，一个住在贫穷社区的聋女，竟然还能当科学家。



但这个人口普查员不知道的是，他嘲笑的这个聋女，不但是一位科学家，还是人类历史上最伟大的女科学家，没有之一。

这个聋女叫亨丽爱塔·勒维特（Henrietta　Leavitt）。她就是本节课一开头那个问题的答案。

本书的其他章节都会同时介绍好几位科学家。因为科学的进步，

往往都是由多位科学家共同推动的。但本章节是个特例。因为在本章节中，只有勒维特一个人的故事。

勒维特的故事得从一场灾难讲起。1892年，刚刚大学毕业的勒维特，按照美国当时的传统，坐船前往欧洲，开始了自己的毕业旅行。

可惜天有不测风云。在这场旅行中，一场突如其来的大病让她的视力和听力严重受损。虽然她的视力后来得到了好转，但是她的听力却每况愈下，直至最终失聪。在此后近30年的时间里，她一直处于体弱多病的状态。

由于这场大病，旅行归来的勒维特无法找到合适的工作。幸好，她拿到了自己本科学院一位教授的推荐信，成了哈佛大学天文系的一名硕士研究生。而她的研究生导师，是哈佛大学天文台台长爱德华·皮克林（Edward Pickering）。



皮克林当时在做一个大项目。他想搞清楚天上的恒星到底有哪些种类。为此，他专门组建了一支完全由女性构成的研究团队（先后有80多位女性加入过这个团队）。她们被后人称为哈佛计算员。

1893年，勒维特也成了哈佛计算员中的一员。





不幸的是，勒维特糟糕的健康状况严重拖累了她的学业。由于体弱多病，她隔三岔五就得请病假，这让她的研究工作变得支离破碎，总是无法完成皮克林布置的任务。在苦苦煎熬了3年以后，1896年，意识到自己已经不可能完成学业的勒维特选择了放弃。她离开了哈佛大学天文台，这一走就是6年。

这6年内发生了什么，我们已经无从得知了。我们只知道，1902年，勒维特给皮克林写了一封信，说自己的处境很艰难：失聪的缘故，她已经找不到其他工作了。因此，她恳求皮克林，让自己重回哈佛大学天文台。皮克林还算好心人，答应了。

但是这回，皮克林多了一个心眼儿。他觉得，体弱多病的勒维特肯定会拖慢自己团队的研究进度。所以，他就没让勒维特参与最重要的恒星分类工作，而派她一个人去研究造父变星。

造父变星是一种非常特殊的恒星。它能像心跳似的，发生周期性

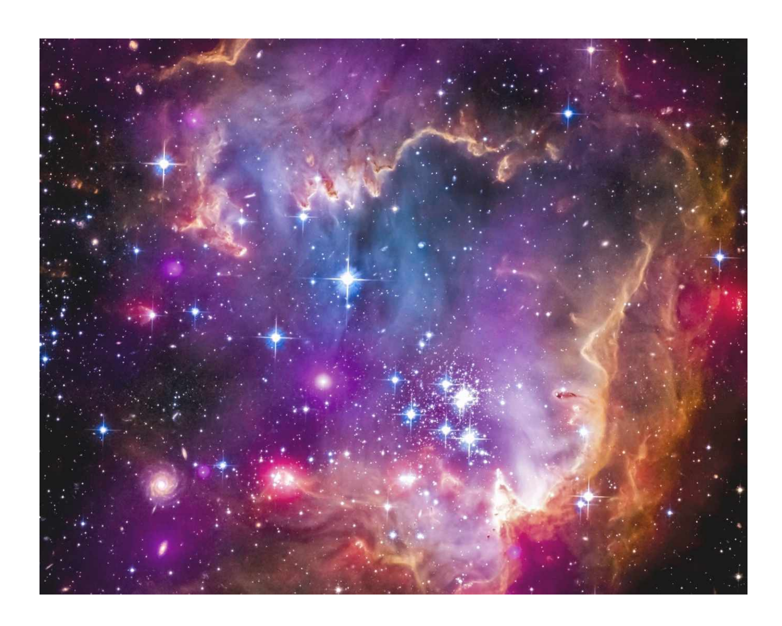
的膨胀收缩，以及周期性的明暗交替。最典型的造父变星是仙王座δ，我国古人管它叫造父一。这就是它中文名字的由来。



在20世纪初，人类连最简单的天上恒星的种类都搞不清楚，就更别提异常复杂的造父变星了。所以在那个年代，派一个人单枪匹马地去研究造父变星，无异于学术上的发配边疆。

现在，让我们暂停一下，来回顾勒维特前半生的电影：由于体弱多病、身心俱疲，她被迫放弃硕士学业，一走就是6年；然后，又由于双耳失聪、家道中落，她迫于生计，不得不重返伤心地；最后，她受到老板嫌弃，直接被发配边疆。

为什么要在这里暂停？因为这是我们与平凡女子勒维特的最后一面。电影重启之后，她将王者归来。



谁也没有料到，勒维特竟是一个搜寻造父变星的顶尖高手。从1904年开始，她就以令人瞠目结舌的速度不断发现新的造父变星。以至于有天文学家专门致信皮克林："勒维特小姐是寻找变星的高手，我们甚至来不及记录她的新发现。"

1908年，勒维特发表了一篇论文，宣布自己在麦哲伦星云中，找到了1777颗造父变星（在此之前，人们找到的造父变星总数只有几十颗）。这个惊人的数字立刻引起了轰动。《华盛顿邮报》还专门对此进行了报道。

但这个惊人的数字及《华盛顿邮报》的报道，并不是我要给你讲勒维特故事的理由。

我之所以要给你讲勒维特的故事，是因为她在这篇论文的结尾，选了16颗位于小麦哲伦星云的造父变星，并用一张表格列出了它们的光变周期（造父变星完成一轮明暗交替的时间）和亮度。此外，她还留下了一句评论："值得关注的是，造父变星越亮，其光变周期就越长。"

4年后，也就是1912年，勒维特又发表了一篇论文，对此结论进行了完善（之所以拖了4年才发表论文，是因为她在此期间又大病一

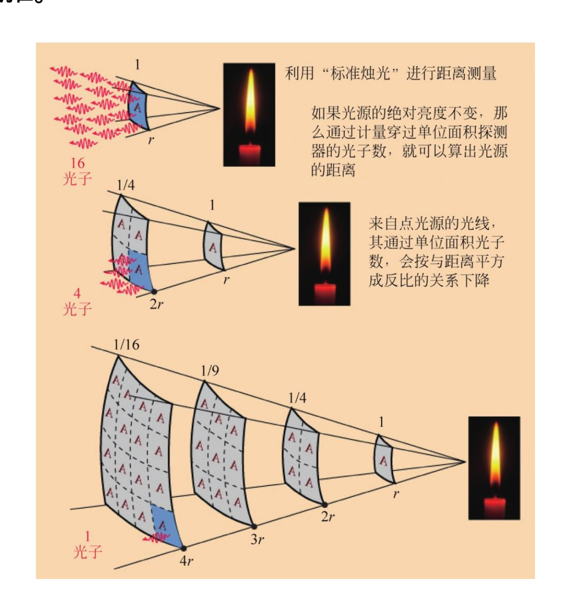
场）。她选了25颗位于小麦哲伦星云的造父变星，把它们画在了一张以亮度为X轴、以光变周期为Y轴的图上。结果表明，这25颗造父变星恰好能排成一条直线。勒维特据此断言：“造父变星的亮度与其光变周期成正比。”

为了理解这句看似平淡无奇的话在天文史上的分量，你不妨想象一片被冰封了不知多少岁月的荒原，由于这句蕴含着巨大魔力的咒语，在转瞬之间就绽放出数以亿计的美丽花朵。

这句话后来被称为勒维特定律。它开创了一个全新的学科——现代宇宙学。

你可能会觉得奇怪了：“为什么这么简单的一句话，能开创一个全新的学科呢？”答案是，它提供了一种全新的天文距离测量方法，那就是著名的“标准烛光”。

为了科普标准烛光，让我们从一个日常生活中很常见的现象说起。一根蜡烛，放在近处看就亮，放在远处看就暗。那么，蜡烛的亮度和我们与它的距离之间，到底有什么关系呢？答案是，亮度与距离的平方成反比。比如说，我们原本与蜡烛相距1米。如果退到2米处，亮度就会变成原来的1/4；如果退到3米处，亮度就会变成原来的1/9。依此类推。



更重要的是，这个数学关系可以反着用。例如，我们站在一座山峰上，想测量它和另一座山峰之间的距离。在两座山峰之间有一个大裂谷，根本就无法过去。那么，怎么才能测出两者间的距离？

答案是，可以另找一个人，让他拿着一根蜡烛，爬上对面的山峰。然后，我们再观测他手中蜡烛的亮度。如果亮度降为了原来的100万分之一，就说明两山之间相距1000米；如果亮度降为了原来的1亿分之一，就说明两山之间相距10000米。

这意味着，蜡烛可以作为一种测量距离的工具。

这个用蜡烛来测量距离的原理，同样可以用到天上。但是，一个天体要想被当成蜡烛，必须同时满足以下两个条件：（1）它必须特别明亮，即使相距甚远也能看到；（2）它自身的亮度必须始终保持不变。这种能当蜡烛用的特殊天体，就是所谓的标准烛光。

问题在于：能同时满足这两个条件、并被视为标准烛光的天体，实在是太少了。

知道了什么是标准烛光，我们就可以介绍勒维特的科学贡献了。





勒维特定律说的是，造父变星的亮度与其光变周期成正比（由于勒维特选出的那些造父变星全都位于小麦哲伦星云内，可以认为它们与地球的距离都相等。这样一来，就不用再考虑距离的因素了）。这意味着，只要选择一批光变周期完全相同的造父变星，就可以得到一批自身亮度完全相同的天体。

这意味着，造父变星可以同时满足标准烛光的两大条件。所以，勒维特真正的发现是，造父变星是一种真正意义上的标准烛光，能够用于天文距离测量。它也是人类历史上发现的第一种标准烛光。

标准烛光的发现，为人类提供了一种全新的测量遥远天文距离的方法。它把人类的天文测距能力，从20世纪以前的几千光年，直接提升到了几亿光年。

由此，人类突破了银河系的禁锢，把目光投向了整个宇宙。现代宇宙学也随之诞生。

所以，勒维特被后世称为“现代宇宙学之母”。她也是目前为止唯

——一个能被称为某个大学科之母的人。

悲哀的是，勒维特的故事并没有一个圆满的结局。

标准烛光的发现，让皮克林意识到了勒维特的厉害。所以，他就给勒维特安排了一份新工作：研究北极星序。简单地说，就是去观测北极星附近的96颗恒星，然后对它们进行分类。

以今天的眼光来看，这个安排可谓荒唐透顶：相当于强迫正值当打之年的飞人乔丹放弃自己的篮球生涯，去参加一个业余的棒球联赛。它让全世界对于变星测光的研究，倒退了至少20年。

而讽刺的是，尽管以一己之力开创了一门后来养活了数万名科研人员的全新学科，勒维特却完全没得到任何世俗意义上的奖励。没有公开表彰，没有教授职位，甚至没有博士文凭。从始至终，她一直是一个薪水只有男人一半的普通计算员。

1921年，勒维特又病倒了。这次，她患上了无药可救的癌症。同年12月12日，勒维特在一个雨夜中离去。她留下遗嘱，把剩下的所有财产都留给了与自己相依为命的母亲。这些遗产包括3张债券和一些家具，价值合计为315美元，还不够买8条地毯。

去世后的勒维特，被葬在了自己家族的墓地。由于贫穷，她甚至无法拥有一个单独的墓碑，被迫和好几个亲戚挤在一起（见图7）。这个墓碑很小，位置只够写她的姓名、生日和忌日。

这就是标准烛光的发现者、哥白尼日心说的掘墓人、“现代宇宙学之母”、人类历史上最伟大的女科学家最后的结局。

100多年过去了，现在勒维特这个名字已经快被世人遗忘在历史的尘埃里了。但我依然想写一篇悼文，来纪念这位非凡女性所经历的种种苦难和荣耀。尽管经历了病痛、失聪、贫穷、孤独、被摆布、被轻视、被遗忘，她依然是照亮整个宇宙的永世不灭的烛火。



图7



勒维特凭一己之力提出了一种全新的天文距离测量方法，即标准烛光。标准烛光的出现，让人类的天文测距能力，从20世纪前的几千光年，一下提升到了几亿光年。正是利用标准烛光人类才发现，银河系并不是宇宙的全部，而是宇宙中的一个小小的孤岛。

那么，人类是如何发现银河系只是宇宙中的小小孤岛的呢？

欲知详情，请听下回分解。

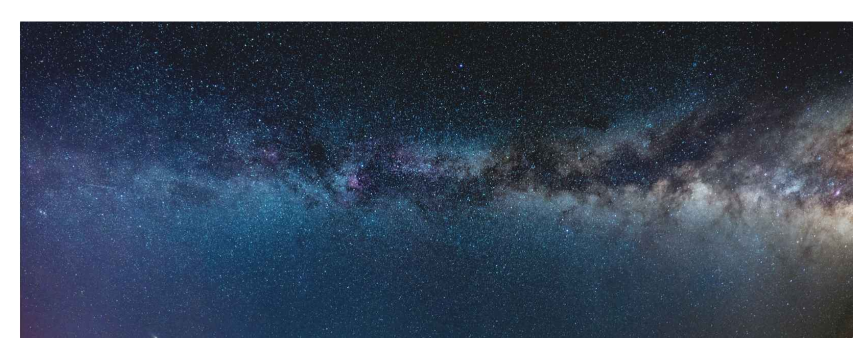
# 4 银河系的大小

上节课的结尾，我们提出了这样一个问题：人类是如何发现银河系只是宇宙中的小小孤岛的呢？

这个问题的背后，同样有一段颇为曲折的历史。

第一个登场的历史人物，是美国著名天文学家哈罗·沙普利（Harlow Shapley）。

沙普利家境贫寒，刚读完小学就被迫辍学。为了谋生，他打过很多零工，包括给一家乡村小报当记者。后来，他重返校园以完成中学学业，并被密苏里大学录取。





由于当过记者，沙普利决定要攻读新闻学的学士学位。但是入学

那天他惊愕地发现，密苏里大学新闻学院当年根本不招生。对其他专业一无所知的沙普利，决定按照英文字母的顺序来选择专业。他放弃了第一个专业archaeology（考古学），因为他读不准这个单词的音。所以，他就选择了第二个专业astronomy（天文学）。

本科毕业后，沙普利考上了普林斯顿大学的博士研究生，师从普林斯顿大学天文系第一位系主任、美国著名天文学家亨利·罗素（Henry Russell）。读博期间，他知道了勒维特提出的标准烛光的概念，并为此深深着迷。

1914年，博士毕业的沙普利被聘为威尔逊山天文台研究员。在那里，他以喜欢蚂蚁而出名，因为他把业余时间都用来观察在混凝土墙上爬行的蚂蚁了（沙普利发现，蚂蚁的爬行速度与外界温度关系密切，所以可以利用蚂蚁来作温度计）。

正所谓时势造英雄。这个喜欢蚂蚁的农家少年，正好赶上了天时、地利、人和。

天时是指，勒维特在两年前发现造父变星是一种标准烛光，可以用来测量遥远的天文距离。

地利是指，沙普利所在的威尔逊山天文台，有一台1.5米口径的光学望远镜。当时，这是全世界最大、最先进的望远镜。

人和是指，沙普利得到了威尔逊山天文台台长乔治·海耳（George Hale）的大力支持。

有了天时、地利、人和，沙普利踌躇满志。他决定利用这台1.5米口径的望远镜，来探测银河系的结构。

银河系有好几千亿颗恒星，一颗一颗地数显然是不现实的。所以，沙普利决定要探测银河系的“骨架”，即球状星团。

球状星团是由几万颗到几百万颗恒星构成的一种非常密集的球状恒星集团（见图8）。目前，人类已经在银河系中发现了100多个球状星团。



图8

沙普利相信，球状星团是银河系的“骨架”，能反应银河系的大小和形状。因此，只要想办法测出这100多个球状星团到地球的距离，便能绘制出整个银河系。

沙普利的测距工作，可以大致分为三个阶段。

第一阶段，在离地球最近的球状星团中，寻找造父变星。只要能找到造父变星，就可以利用它们来确定最近的球状星团的距离。

第二阶段，在找不到造父变星的较远的球状星团中，寻找另一类变星，即“星团变星”。然后把星团变星也视为标准烛光(一种天体要是能被视为标准烛光，就可以利用其亮度与距离的平方成反比的关系，来进行天文测距)，并用同时拥有造父变星和星团变星的球状星团进行定标。这样就可以用星团变星作量天尺，来确定较远的球状星团的距离。

第三阶段，对于那些什么变星都找不到的球状星团，直接把整个球状星团都视为标准烛光；然后再用相距较近且距离已知的球状星团，来估算那些遥远球状星团的距离。

需要特别强调的是，沙普利第二阶段和第三阶段的距离测量其实并不准确。原因在于，无论是星团变星还是球状星团，都不是真正的标准烛光；换句话说，用星团变星或球状星团来测距，其实并不准确。不过，这些偏差，并不影响沙普利对银河系形状的最终绘制结果。

经过数年的努力，沙普利终于完成了对这100多个球状星团距离的测量，进而绘制出了整个银河系的骨架。他发现，太阳根本就不是银河系的中心；银河系真正的中心在人马座方向，离我们至少上万光年。



正是这个发现，宣判了哥白尼日心说的“死刑”。沙普利也由此名动天下。

但是功成名就的沙普利，并没有逃脱屠龙少年终成恶龙的宿命。没过多久，他就成了一场世纪大辩论的反派人物。

为了更好地介绍这场世纪大辩论，我得先给你补充一些背景知识。

之前已经说过，在20世纪前，人类普遍相信银河系就是宇宙的全部。

不过，也有人反对这幅宇宙图像。他们相信，宇宙是一片浩瀚的大海，而银河系只是漂浮在这片大海上的一座小小的岛屿。在银河系之外，还有许许多多和它一样大小的岛屿。这就是所谓的“宇宙岛”理论。

“宇宙岛”理论的代表人物之一，是德国大哲学家伊曼纽尔·康德（Immanuel Kant）。

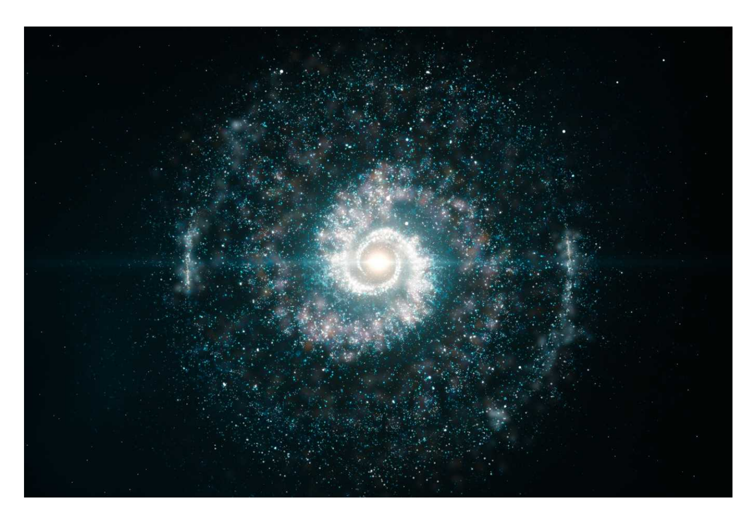


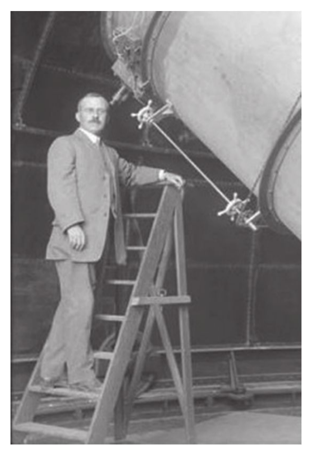
当时天文学家已经在银河系边缘发现了一些螺旋星云，不过一直无法确定它们与地球相距多远。康德就提出了一个大胆的猜想：这些螺旋星云全是和银河系一样大小的岛屿。

但后来，人们发现这样的螺旋星云有不下10万个。对那个年代的天文学家来说，想象在银河系外还有10万个与银河系一样大小的星系，完全是不可理喻。

所以，“宇宙岛”理论就进了天文学界的冷宫，这一关就是100多年。

直到1914年，“宇宙岛”理论才得以东山再起。那一年，美国天文学家维斯托·斯里弗（Vesto Slipher）想到了一个好办法，能精确测出螺旋星云的运动速度（斯里弗的办法是本书第6节课的核心内容，到时会详细介绍）。他测量了15个螺旋星云，发现速度最快的两个星云，正以1100千米/秒的超高速飞离地球。很难想象，银河系内部的天体能达到如此恐怖的速度。



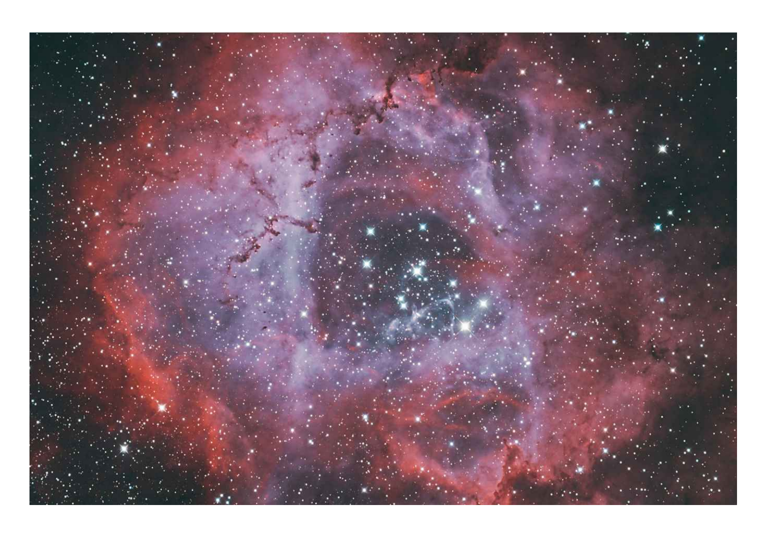


过了3年，又出现了支持“宇宙岛”理论的新证据。美国天文学家赫伯·柯蒂斯（Heber　Curtis）在螺旋星云中找到了不少新星（新星就是突然出现在天空的明亮星星，我国古代称为客星）。他把新星视为标准烛光，利用标准烛光亮度与距离的平方成反比的关系，对这些螺旋星云的距离进行了测量。柯蒂斯的测量结果显示：这些螺旋星云与地球相距甚远，远超当时公认的银河系直径，即3万光年。

但这些发现并没有让“宇宙岛”理论成为天文学界的主流。其中最大的障碍，就是在本节课中最早出场的沙普利。他在绘制银河系的时候算出，银河系的直径能达到惊人的30万光年。所以他认为，这么大的空间区域，足以装下整个宇宙。

讲完了背景知识，接下来我们就可以聊聊天文学的世纪大辩论了。

1920年年初，为了提升自身的影响力，美国科学院决定要找两个大牌科学家，做一场面向普通民众的公开辩论。美国科学院秘书提议，可以搞一场关于冰川的辩论。



这个提议遭到了一个实权派人物的极力反对，此人就是威尔逊山天文台台长乔治·海耳（George Hale）。

海耳认为，要想产生较大的社会影响力，就应该选择那些最前沿的科学辩题。所以，他推荐了两个新的辩题。



第一个辩题是爱因斯坦的相对论。此辩题让美国科学院秘书气得吐血。他宣称：“应该把相对论扔到四维时空以外的某个地方，这样它就不会再来困扰我们了。”

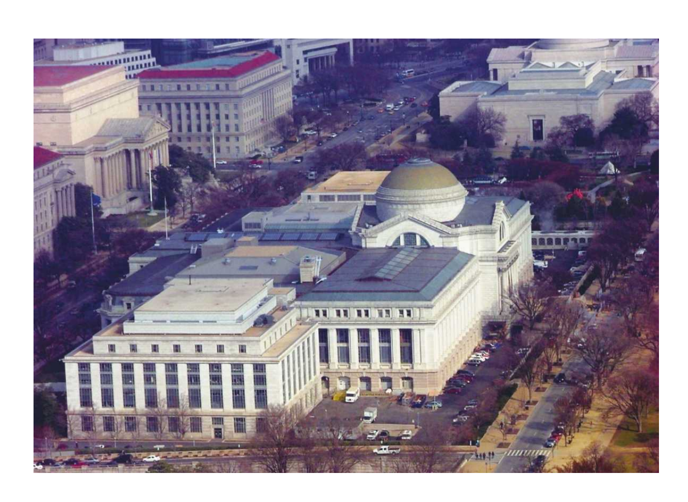
所以最后就采用了海耳推荐的第二个辩题，即宇宙的尺度。

有了辩题，接下来就该选择辩手了。根据海耳的建议，最后确定了两个最合适的人选，也就是之前介绍过的沙普利和柯蒂斯。

大辩论的正方是沙普利。他持传统观点，认为银河系就是宇宙的全部；而反方是柯蒂斯。他相信“宇宙岛”理论，认为银河系外还有很多其他的星系。而双方争论的焦点是，银河系到底有多大。

1920年4月26日，这场举世瞩目的世纪大辩论，在纽约市史密斯森自然历史博物馆拉开了帷幕。

辩论开始后，沙普利首先出场。但他还没开口说话，就已经落了下风。这是因为，此时的他已经开始患得患失了。



不久前，哈佛大学天文台台长皮克林因病去世；很快，哈佛大学就开始甄选下一任台长。沙普利已经从秘密渠道得知，自己就是这个职位的热门人选。所以，他很害怕自己输掉这场辩论，给坐在台下的哈佛大学代表留下坏印象。

因此，沙普利彻底改变了原来的辩论策略。在描述如何测量银河系大小的时候，他直接跳过了自己实际采用的造父变星和球状星团，

而介绍了自己根本没用的蓝巨星（蓝巨星是一种非常明亮、非常炽热的恒星）。沙普利宣称，蓝巨星也是一种标准烛光，可以用其来测量银河系的大小；而蓝巨星的测量结果表明，银河系的直径能达到惊人的30万光年，这足以容纳整个宇宙。

其实，蓝巨星仅仅是沙普利为这场辩论准备的脚注材料。

沙普利这么做的理由很简单。他用造父变星和球状星团进行天文测距的工作，在学术界早已尽人皆知。如果在辩论时介绍这两种测距工具，肯定会遭到柯蒂斯有针对性的攻击。所以，他干脆介绍柯蒂斯不知道的蓝巨星，这样还能打对手一个措手不及。

很明显，这是典型的未战先怯。

沙普利的诡辩策略并没有打乱柯蒂斯的阵脚。他很快就向沙普利发起了强大的攻势。

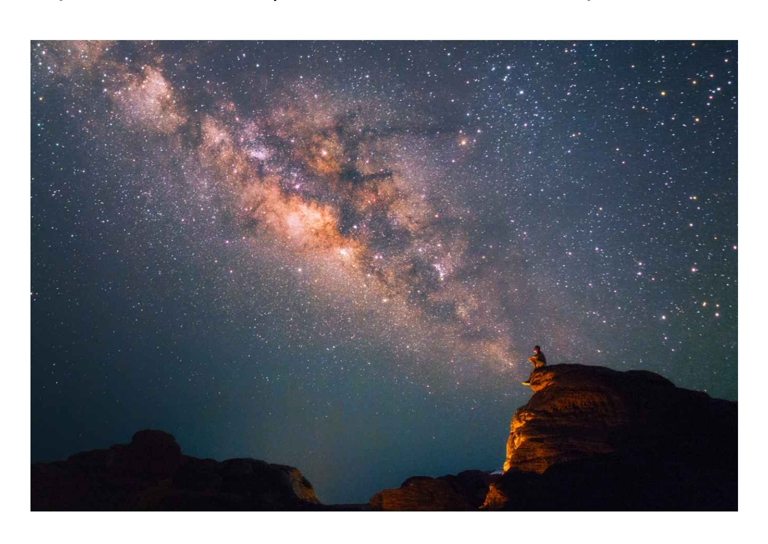
首先，柯蒂斯质疑了沙普利采用的测距工具的可靠性；无论是造父变星、球状星团，还是蓝巨星，都没逃过他的批驳。其次，他提出了一种他认为更可靠的测距工具，即黄矮星（太阳就属于黄矮星）。接着，把黄矮星视为标准烛光，柯蒂斯估算了银河系的直径，结果只有3万光年。最后，柯蒂斯描述了“宇宙岛”的物理图像，并解释了螺旋星云为何只出现在银河系的上下两端。



毫无疑问，柯蒂斯成了这场大辩论的赢家。在给家人的信中，他颇为得意地写道："华盛顿的论战很顺利，我确信我更胜一筹。"没过多久，他就被聘为阿利刚天文台台长。

沙普利的日子就没这么好过了。辩论会上的糟糕表现，让他差点丢掉了哈佛大学天文台台长的位子。因为哈佛大学实在找不到其他的合适人选，沙普利最后还是如愿以偿地成了皮克林的继任者。

这场世纪大辩论，让"银河系到底有多大"的学术论战，进入了公众的视野。不过搞笑的是，辩论双方其实都是错的。



沙普利基于球状星团测出，银河系直径是30万光年；柯蒂斯基于黄矮星，测出银河系直径是3万光年。而目前的天文观测结果表明，银河系直径的实际值应为10万光年。

为什么这两个顶级天文学家都搞错了呢？答案是，他们的距离测量工具有问题。无论是球状星团还是黄矮星，都不是能用来测距的标准烛光。真正靠谱的测距工具，还是勒维特提出的那个造父变星。

这场世纪大辩论并没有改变天文学界的分裂格局。对于螺旋星云是否属于银河系，依然是公说公有理，婆说婆有理。

直到3年后，一个年轻人的横空出世，才为这场世纪大辩论画上了句号。此人就是美国大天文学家埃德温·哈勃（Edwlin Hubble）。

与幼年辍学、历尽磨难的沙普利不同，学生时代的哈勃，可谓一路顺风顺水。



读高中的时候，身高1.9米的哈勃是一个不折不扣的明星运动员，曾在一次市级的中学生运动会上，一口气拿到了7个冠军。靠着运动特长，他被保送到了芝加哥大学；而在毕业前夕，他又拿到了罗德奖学金，得以前往英国牛津大学攻读法学硕士学位。

学成归国的哈勃，并没有从事法律工作，因为他没能通过美国的司法考试。无奈之下，他只好跑到家乡的一所高中任教，主讲数学，同时兼任校篮球队教练。所以，谁说体育老师不能教数学。

后来，在一位本科教授的帮助下，哈勃得以重返芝加哥大学，攻读天文学博士学位。第一次世界大战后，博士毕业且服完兵役的哈勃，被聘为威尔逊山天文台研究员。在那里，他遇到了自己一生的宿敌，那就是我们的老朋友沙普利。

一山不容二虎。沙普利从一开始就看不惯这个总是一身英伦范（穿短裤、顶斗篷、叼烟斗）、却没有做出什么成绩的年轻人。由于那时的沙普利在美国天文学界如日中天，哈勃最初在威尔逊山天文台工作的日子并不好过。

但没过多久，哈勃的职业生涯就迎来了转机。他任职的威尔逊山天文台新建了一个2.5米口径的“胡克望远镜”；而在此后30多年的时间里，这个“胡克望远镜”一直是全世界最大、最先进的光学望远镜。与此同时，沙普利离开了威尔逊山天文台，接任哈佛大学天文台台长。这样一来，哈勃就得到了大量的、本属于沙普利的“胡克望远

镜”观测时间。

坐拥全世界最大望远镜之上的广袤天际，运动员出身的哈勃即将一飞冲天。

1923年，哈勃利用“胡克望远镜”，在仙女星云中发现了两颗造父变星（毫无疑问，造父变星是最靠谱的标准烛光）。利用这两颗造父变星，哈勃测出仙女星云与我们相距至少100万光年。这个惊人的数字已经远远超越了银河系的尺寸，说明仙女星云必然位于银河系之外。

哈勃把这个发现写成了一封信，用急件寄给了沙普利。根据沙普利的高徒、哈佛大学历史上首位女系主任塞西莉亚·佩恩（Cecilia Payne）的回忆，沙普利收到哈勃的信后如雷轰顶，哀叹道：“这封信摧毁了我的世界。”

当然，沙普利并不甘心坐以待毙。数日后，他给哈勃写了封回信，质疑哈勃在仙女星云中找到的那两颗造父变星都是“伪星”，根本不能用于天文学测距。

但沙普利最后的挣扎，已经无法阻挡历史的车轮滚滚向前。

1924年，哈勃继续观测星空，并取得了新的进展。他成功地在仙女星云、巴纳德星云以及M33星云中找到了更多的造父变星。无一例外地，它们全都揭示出银河系并非宇宙的全部。

哈勃的发现，让这场从1920年开始的世纪大辩论彻底落下了帷幕。天文学界达成了共识：人类长期相信是整个宇宙的银河系，其实仅仅是“星云王国中的一个小小的村落”。

我们已经介绍了人类探索银河系大小的曲折历史。现在我们知道，形如圆盘、直径10万光年的银河系，仅仅是浩瀚宇宙中的一个小小的孤岛。

那么，整个宇宙到底有多大呢？

欲知详情，请听下回分解。



# 5　可观测宇宙的大小

上节课的结尾，我们提出了这样一个问题：整个宇宙到底有多大?为了给你一个更直观的感受，我将用比喻的方式来回答这个问题。

如果把太阳系想象成一栋别墅，那么地球就是这栋别墅里的一颗玻璃珠。

4000亿栋“别墅”合在一起，构成了一个“中心城区”。这个“中心城区”叫银河系。

银河系和另外一个“中心城区”（即仙女座星系），再加上周边的50个小型星系，就构成了一座“城市”。这座“城市”叫本星系群。

本星系群只是一座“小城”。在离它5000万光年远的地方还有一座拥有2000个星系的“大城市”，叫室女座星系团。以这个室女座星系团为“省会”，再加上方圆1亿光年内的100多个“城市”，就构成了一个“省”。这个“省”叫室女座超星系团。



室女座超星系团只是4个“省”之一。这4个“省”像群山一样，环绕着一个位于中心谷地处的“首都”（即巨引源，与地球相距2.2亿光年，其质量能达到银河系的10000倍）。这样一来，就在直径5亿光年的空间范围内，构成了一个地形如同巨大山谷的“国家”。这个“国家”叫拉尼亚凯亚超星系团。

拉尼亚凯亚超星系团并不是一个“大国”。它连同周边的4个“国家”，构成了一个“国家联盟”，叫双鱼-鲸鱼座超星系团复合体。此“国家联盟”的盟主是双鱼-鲸鱼超星系团，其疆域至少能达到10亿光年。

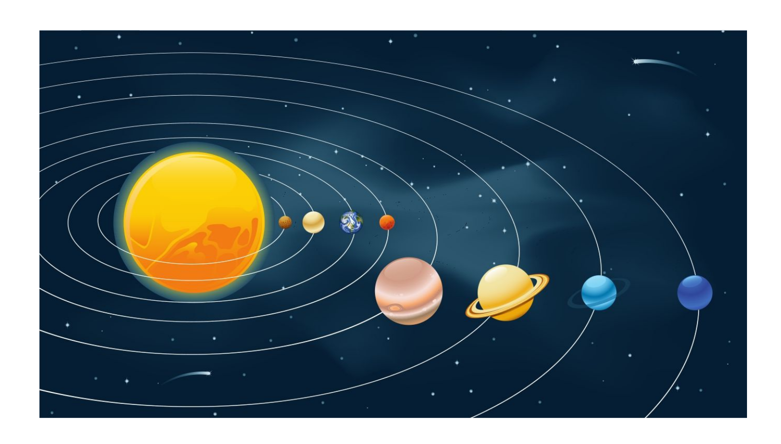
横跨10亿光年的双鱼-鲸鱼座超星系团复合体，依然不是宇宙中最大的结构。在它之上还有所谓的星系长城，相当于“大洲”。比较有名的“大洲”包括横跨14亿光年的史隆长城，横跨40亿光年的巨型超大类星体群，以及横跨100亿光年的武仙-北冕座长城。而这个与地球相距100亿光年的武仙-北冕座长城，就是人类目前所发现的最大结构。

而诸多“大洲”又构成了一个直径930亿光年的“星球”。这个“星球”就是我们的可观测宇宙（所谓的可观测宇宙，是指以地球为中心、用望远镜能够看到的最大宇宙范围。它只是整个宇宙的一小部分）。

我们已经对宇宙的大小做了一个简要的介绍。你可以把我们能够看到的宇宙，想象成一个直径930亿光年、拥有至少几千亿个星系的巨大“星球”。那接下来，就让我们坐上宇宙飞船，按由近及远、由小到大的顺序，来好好看看这个浩瀚的宇宙。

首先参观我们住的这个“别墅”，即太阳系。

太阳系的绝对主宰是太阳。其质量能达到200万亿亿亿千克，是地球的33万倍，占太阳系总质量的99.86%。正因为如此，太阳系内的所有其他天体都必须臣服于它，并且周而复始地围绕它旋转。

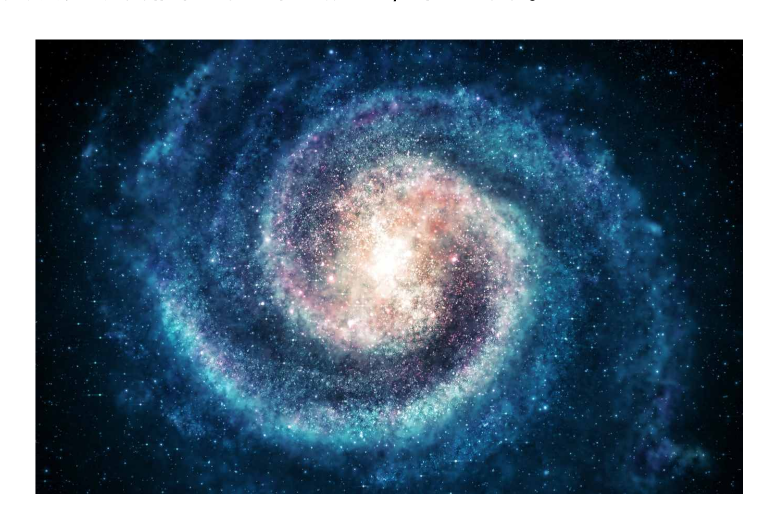


太阳系总共有8环：从内到外，依次是水星、金星、地球、火星、木星、土星、天王星和海王星。值得一提的是，里面的4个都是岩质行星，也就是以硅酸盐岩石为主要成分的行星；在岩质行星的中心，一般都有一个以铁为主的金属内核。而外面的4个都是气态行星，也就是最外层区域由气体构成的行星；其中木星和土星的外层主要成分是氢气和氦气，而天王星和海王星的外层则包含大量比氢和氦更重的化学元素；所有气态行星的中心，同样有由岩石和金属所构成的坚固内核。

此外，太阳系内还有两个小行星聚集的区域，一个是位于4环和5环间的小行星带，另一个是位于8环以外的柯伊伯带。曾是太阳系第9大行星的冥王星，就位于这个柯伊伯带。在柯伊伯带之外，还有一个包裹着太阳系、直径约为2光年的神秘球状云团，叫作奥尔特云，它是很多长周期彗星的故乡。而这个奥尔特云，就是太阳系“别墅”的外

墙。

其次参观我们住的这个“城区”，即银河系。



在银河系的正中心，盘踞着一个质量能达到太阳质量400多万倍的巨大黑洞，叫人马座A*。在它的周围有一个恒星相当密集的棒状区域，其长度约为1万光年（因为中心区域是棒状的，银河系被称为棒旋星系）；这个棒状区域是一个巨大的恒星“育婴室”，其中包含着大量的新生恒星。中心黑洞和棒状区域，统称为银心。

在银心之外，是一个直径接近10万光年的盘状结构，称为银盘。正如图9所示，在银盘上有几个恒星比较密集的区域：其核心特征是从银心附近出发，螺旋式地向外延展。这些银盘上的恒星密集区域就是所谓的旋臂。银河系的主要旋臂有4条：其中青色的是三千秒差距-英仙旋臂，紫色的是矩尺-天鹅旋臂，绿色的是盾牌-半人马旋臂，红色的是船底-人马旋臂（图9中的虚线代表这些旋臂在理论上存在、但实际上尚未观测到的部分）。除了这4条主要旋臂以外，还有一些次要旋臂，例如，图9中橙色的猎户旋臂。我们住的太阳系“别墅”，就位于这个猎户旋臂中。换句话说，我们其实住在银河系“城区”中一个比较荒凉的地段。

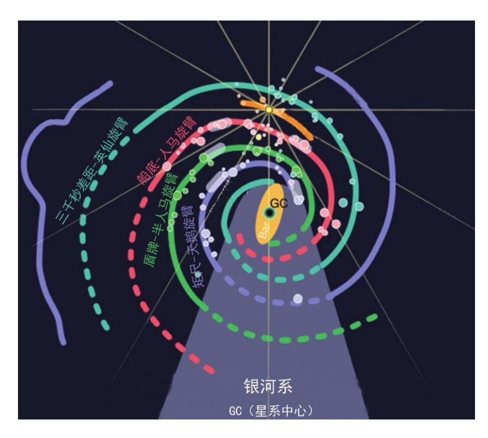



图9

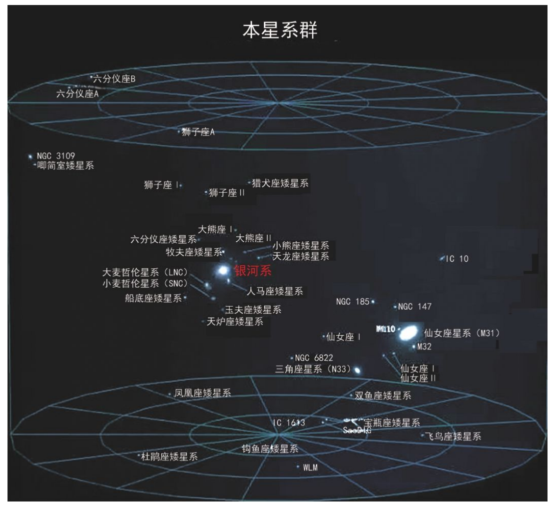
银河系的旋臂，就像风车一样不断绕银河系中心旋转。需要强调的是，旋臂里的恒星构成并非固定不变。为了便于理解，你可以把旋臂想象成银河系“城区”的交通拥堵区域。因为不断有恒星进入这些区域，同时又不断有恒星离开，所以总体来看，这些交通拥堵区域会一直存在。恒星拥堵区域，这就是旋臂的本来面目。

而在银盘之外还有一个球状区域，称为银晕。不同于银心和银盘，银晕内部只是稀稀落落地分布着一些非常古老的恒星和球状星团。你可以把银晕当成是银河系的恒星“养老院”。

由银心、银盘和银晕这3大部分构成的、直径约为10万光年的空间区域，就是我们住的银河系“城区”（也有一些天文学家主张，在银晕之外还有银冕，这样银河系的直径能达到将近20万光年）。

然后参观我们住的“城市”，即本星系群。






在本星系群中有两个中心城区，也就是银河系和仙女座星系。而这两大中心城区还各有一群小弟，也就是所谓的矮星系[(1)]。

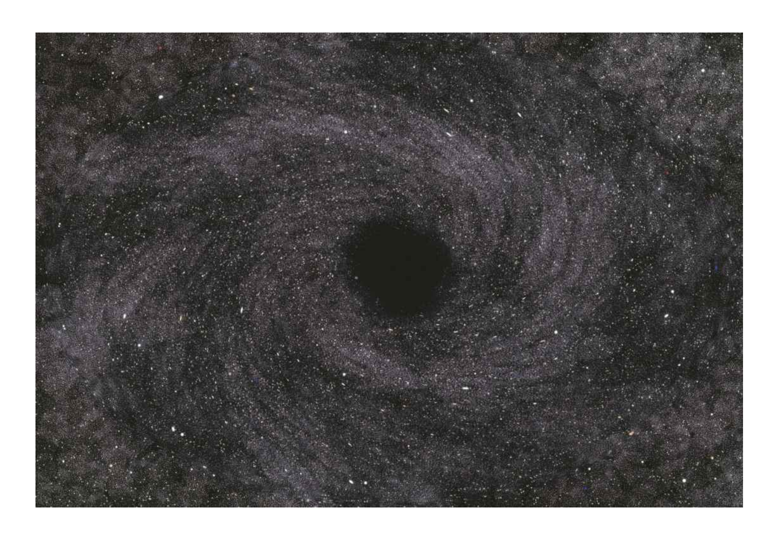
我们居住的银河系，大概拥有4000亿颗恒星；而在它的周围，大概有30多个矮星系。其中离地球最近的是大犬座矮星系和人马座矮椭球星系，而名气最大的则是大小麦哲伦星云。大部分矮星系就像卫星一样绕银河系公转，因而也被称为卫星星系。其他矮星系则只是从银河系周围飞掠而过。

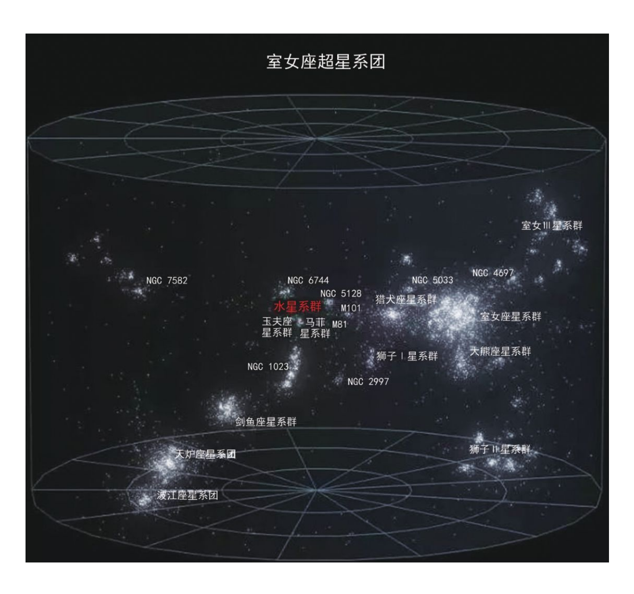
另一个中心城区，是与地球相距254万光年的一个旋涡星系，即

仙女座星系。仙女座星系是本星系群中无可争议的老大，其直径能达到22万光年，而质量能达到太阳质量的1.5万亿倍。类似于银河系，在仙女座星系中心，同样盘踞着一个超大质量的黑洞，其质量能达到太阳质量的1亿倍，是银河系中心黑洞质量的20多倍。

仙女座星系之所以能成为本星系群的霸主，是因为它已经吞并了大量的矮星系。比较有名的例子，是位于仙女座星系内部的一个非常巨大的球状星团，编号G1。一般的球状星团包含的恒星数量，都在几百个到几万个之间；而G1包含的恒星数量，能达到好几百万个。所以天文学家普遍相信：G1是一个矮星系被仙女座星系吞并后，所剩下的致密核心。

最惊悚的是，天文观测表明，仙女座星系正在以每秒110千米的速度，向银河系飞驰而来。大概再过40亿年，两者就会发生碰撞，最终并合成一个巨大的椭圆星系（按照形状，星系可以大致分为3类，分别是旋涡星系、椭圆星系和不规则星系）。



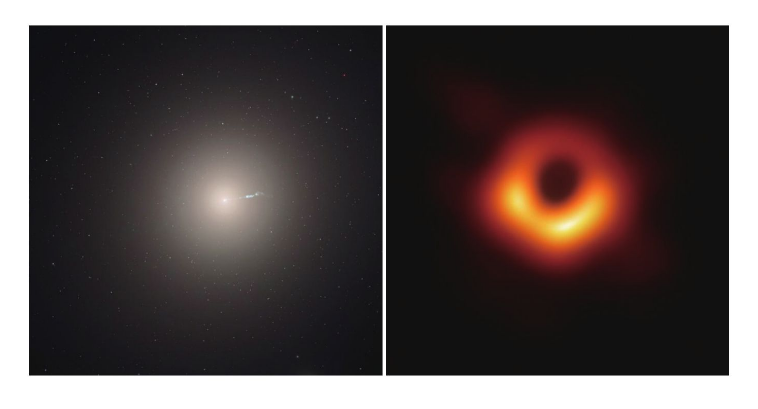



接着参观我们住的“省”，即室女座超星系团。

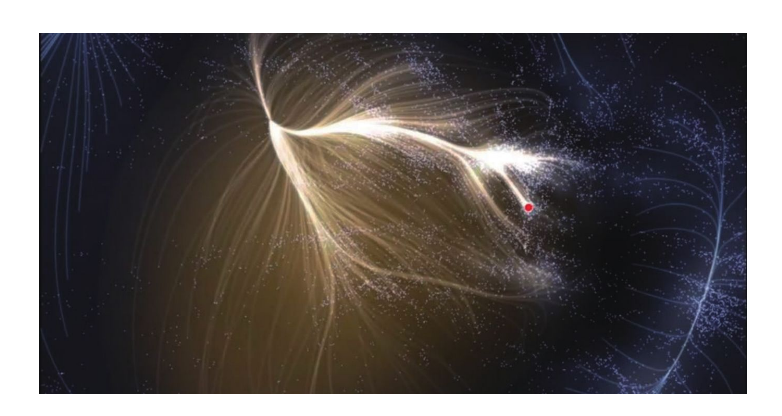
我们住的这个本星系群，只能算一座“小城”。在离它6000万光年远的地方，有一座拥有2000多个星系的“大城市”，叫室女座星系团。

室女座星系团由3个主城区构成，分别是M87、M86和M49（M表示梅西叶星表，而M87、M86和M49分别代表梅西叶星表中的第87、第86和第49号天体）。而这3个主城区，都是超巨椭圆星系。

其中最有分量也最靠近“市中心”的“主城区”，是M87星系。这是一个非常古老的星系，拥有大概15 000个球状星团，堪称恒星的“养老院”。M87星系最显著的特征，是有一条绵延数千光年的星际喷流。此外，在它的中心，有一个质量能达到太阳质量65亿倍的巨型黑洞，叫M87*。2019年，事件视界望远镜项目组拍到了M87*的照片。这也是人类历史上拍到的首张黑洞照片。



由M87、M86和M49这3大主城区构成的室女座星系团，是我们居住的这个“省”的“省会”。这个省还有大概100个“城市”，其中绝大多数都是和本星系群一样的小城，也就是由几十个星系所构成的星系群。只有在这个“省”的边境位置，才有两个中等规模的“城市”，即天炉座星系团和波江座星系团。这100多个“城市”散布在“省会城市”（即室女座星系团）周边直径1.2亿光年的范围内，从而构成了我们住的这个“省”，即室女座超星系团，也叫室女座超星系团。

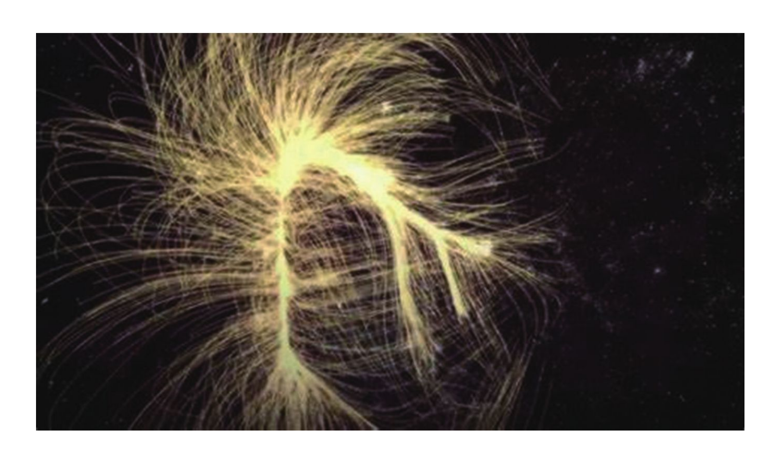


接下来参观我们住的“国家”，即拉尼亚凯亚超星系团。

拉尼亚凯亚超星系团是一个横跨5亿光年、质量能达到银河系质量10万倍的庞大帝国。它的地形有点像是一个巨大的山谷，位于中心谷地位置的就是这个“帝国”的“首都”：巨引源（巨引源代表“巨大的引力源头”）。

巨引源是一个真正意义上的庞然大物，其质量至少比银河系质量

大10000倍。由于它拥有巨大的引力，包括银河系在内的成千上万的星系，都在以每秒几百千米的速度朝它靠近。而这个庞然大物的本来面目到底是什么，目前还处于迷雾中。



在巨引源这个首都的周围还有4个省。处于中心位置的省是长蛇-半人马座超星系团。这个省是在长蛇座到半人马座方向上的一系列星系团的集合。其中最核心的成员是矩尺座超星系团。这个矩尺座超星系团位于矩尺座方向、与地球相距大概2.2亿光年的地方。一般认为，这就是巨引源所在的地方。不过，矩尺座超星系团的质量只比银河系质量大1000倍，仅仅是巨引源质量的1/10。所以，矩尺座超星系团仅仅是占据了巨引源所在的位置，而并非巨引源本身。



除了位于“首都”位置的矩尺座超星系团以外，长蛇-半人马座超星系团这个“省”还拥有长蛇座星系团、半人马座星系团、IC4939星系团这3个“大都市”，以及上百个零零星星的“小城”。这些大大小小的“城市”环绕在巨引源的周围，就构成了拉尼亚凯亚帝国的“首都”圈都市群。

而在“首都”圈的外围，还有3个“省”，分别是位于西南方的室女座超星系团、位于西北方的孔雀-印第安超星系团，以及位于南面的南方超星系团。

为了便于理解，你可以把引力想象成蜘蛛丝。在引力的牵引下，长蛇-半人马座超星系团、室女座超星系团、孔雀-印第安超星系团和南方超星系团这4个“省”，就连成了一张直径5亿光年的巨大蜘蛛网，从而覆盖了拉尼亚凯亚帝国的整个山谷。位于蜘蛛网上的成千上万的星系，都在引力蛛丝的牵引下向着位于中心谷地位置的巨引源运动。这就是我们住的“国家”，即拉尼亚凯亚超星系团的全貌。

我们已经介绍了我们住的“国家”，即拉尼亚凯亚超星系团。它拥有横跨5亿光年的辽阔疆域和超过10万个像银河系这样的星系。但放眼宇宙，拉尼亚凯亚超星系团依然只能算是“弟弟”。在它之上还有由

多个“国家”组成的“国家联盟”，也就是由多个超星系团构成的超星系团复合体。

拉尼亚凯亚超星系团与4个“国家”一起，组成了一个“国家联盟”，叫双鱼-鲸鱼座超星系团复合体。它的疆域能超过10亿光年，并且总质量至少比太阳质量大100亿亿倍。这个“国家联盟”的盟主是双鱼-鲸鱼超星系团，具体成员还包括英仙-双鱼座超星系团、飞马-双鱼座超星系团、玉夫-武仙座超星系团和拉尼亚凯亚超星系团。

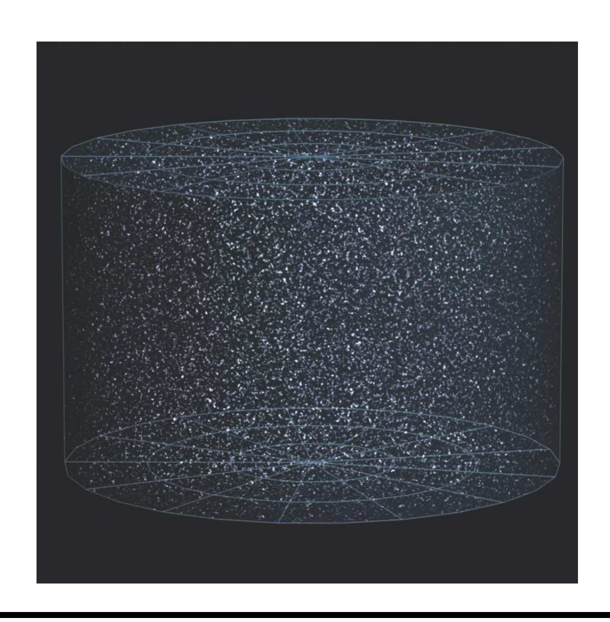
这个横跨10亿光年的双鱼-鲸鱼座超星系团复合体，依然不是宇宙中最大的结构。在它之上还有所谓的星系长城，相当于“大洲”。比较有名的“大洲”包括横跨14亿光年的史隆长城，横跨40亿光年的巨型超大类星体群，以及横跨100亿光年的武仙-北冕座长城（武仙-北冕座长城与地球的距离，大概是100亿光年）。这个武仙-北冕座长城，就是人类目前发现的最大结构。

而诸多的宇宙空洞和星系长城，又构成了一个直径930亿光年的“星球”。这个“星球”就是我们的可观测宇宙。所谓的可观测宇宙，指以地球为中心、用望远镜能够看到的最大宇宙范围。它只是整个宇宙的一小部分。在可观测宇宙之外，还有更辽阔的其他宇宙空间。但其他宇宙空间发生的事情，我们在地球上永远也不可能看到。

我们已经完成了这场飞向宇宙尽头的旅行。现在你应该已经知道，我们能够看到的可观测宇宙，是一个直径930亿光年、拥有至少几千亿个星系的巨大“星球”。在可观测宇宙之外，还有更为辽阔的宇宙空间，但我们永远都不可能看到那里到底有什么东西。

为什么我们永远都不可能看到可观测宇宙之外的宇宙空间呢？

欲知详情，请听下回分解。



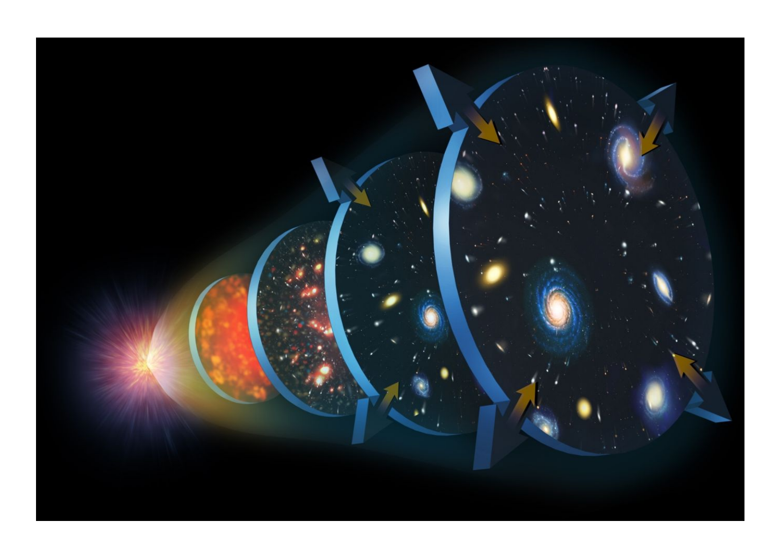
---

(1)矮星系是宇宙中质量最小、亮度最弱的一类星系。不过矮星系的数量远远超过大星系。

# 6　宇宙膨胀

上节课的结尾，我们提出了这样一个问题：为什么我们永远都不可能看到可观测宇宙之外的宇宙空间呢？

答案是，因为宇宙在膨胀。



想象一个巨大的气球，上面有一只小蚂蚁，正以光速在气球表面爬行。如果气球静止不动，那么蚂蚁就能到达气球表面的任意位置；换句话说，蚂蚁能看到气球表面的全貌。但如果气球本身也在以光速膨胀，那么蚂蚁就无法到达气球表面的任意位置了；这意味着，蚂蚁只能看到以其出发点为中心的一小块区域。而蚂蚁能看到的这一小块区域，就是它的“可观测气球表面”。

同样的道理，如果宇宙本身也在膨胀，我们就只能看到以地球为中心的一小块宇宙区域，即可观测宇宙。

那么问题来了：人类到底是如何发现宇宙在膨胀的？

你可能会说：“这还用问吗？宇宙膨胀是美国大天文学家哈勃在20世纪30年代初发现的。”

真实的历史，并没有这么简单。

推倒第一张多米诺骨牌的人，其实并不是哈勃。此人在我们之前的旅行中曾经露过一面。他就是美国天文学家维斯托·斯里弗。

1914年，斯里弗提出了一种测量星系径向速度（即星系与地球连线方向上的速度）的新方法。这种方法的基石是多普勒效应。



什么是多普勒效应？让我们从一个在日常生活中很常见的场景说起。如果你经常坐地铁，可能会注意到这样的现象：当列车进站的时候，它发出的汽笛声会比较尖锐；而当列车出站的时候，它发出的汽笛声会比较低沉。

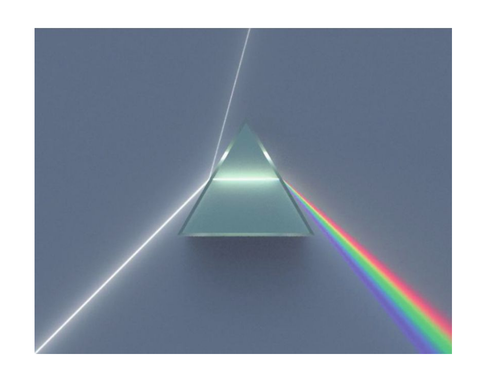
这就是多普勒效应。这个效应说的是：如果一个物体在靠近我们时，它发出的声波波长会变短，频率会变大，所以听起来尖锐；如果一个物体在远离我们时，它发出的声波波长会变长，频率会变小，所以听起来低沉。

多普勒效应的强大之处在于，它不仅适用于声波，还适用于宇宙中所有的波。我们知道，光也是一种波（即电磁波）。那么，多普勒效应会如何影响遥远天体所发出的光呢？

为了回答这个问题，我得先给你补充两个知识点。第一个要补充的知识点，是恒星光谱。



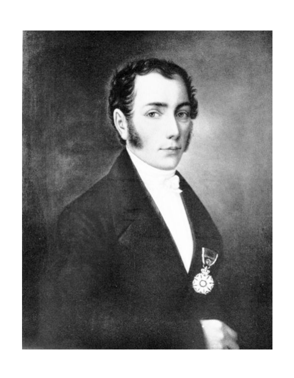
1666年，在乡下躲避瘟疫的艾萨克·牛顿（Isaac Newton）爵士做了一个著名的光学实验，即牛顿色散实验。他在一间小黑屋的墙上开了一个小圆孔，然后在小圆孔的旁边放了一个三棱镜[(1)]。

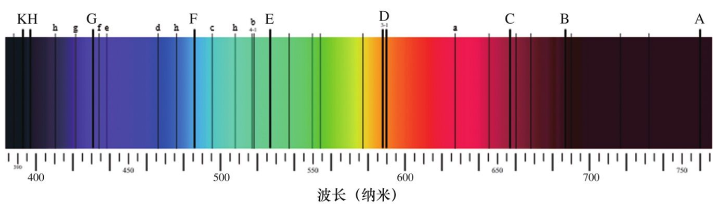


牛顿发现：穿过小圆孔的白色太阳光，在经过三棱镜的折射后，变成了一个由红、橙、黄、绿、蓝、靛、紫等不同颜色构成的光斑。由此，牛顿证明了太阳光并非只有一种单一的颜色，而是由各种颜色的光合成的。

后来，天文学家给望远镜配上了三棱镜。遥远天体发出的光，先透过望远镜的镜面，再经过一条狭缝，然后被三棱镜折射，最后会变成一条让各种单色光按频率大小依次排列的光带。这条让各种单色光按频率大小依次排列的光带，就是所谓的光谱。

第二个要补充的知识点，是夫琅禾费线。

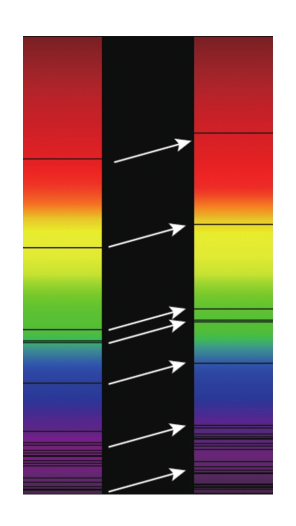





1814年，德国物理学家约瑟夫·夫琅禾费（Joseph　Fraunhofer）用自己发明的新仪器，研究了太阳光谱。他惊讶地发现，在太阳光谱中有超过570多条的暗线。换句话说，有一些特定频率的光从太阳光谱中消失了。这些暗线，就是所谓的夫琅禾费线。

后来科学家发现，之所以会出现夫琅禾费线，是因为太阳表面的化学元素把这些特定频率的光给吸收了（任何一种化学元素，都只能吸收特定频率的光）。因此，只要把天体光谱中的夫琅禾费线与各种化学元素的吸收线（即各种化学元素所吸收的特定频率的光）进行对比，就可以确定天体表面到底有哪些化学元素。

知道了光谱和夫琅禾费线的概念，我们就能介绍多普勒效应是如何影响遥远天体发出的光了。



如果一个天体正在靠近地球，那么在其光谱中，夫琅禾费线就会整体地向蓝光端（即频率变大的方向）移动，这就是所谓的蓝移；而如果一个天体正在远离地球，那么在其光谱中，夫琅禾费线就会整体地向红光端（即频率变小的方向）移动，这就是所谓的红移。

基于遥远天体光谱中的蓝移或红移，就能判断这些天体是在靠近还是在远离地球；而通过测量它们光谱的蓝移或红移的大小，就可以算出这些天体靠近或远离地球的径向速度。

1914年，斯里弗研究了15个随机选取的螺旋星云的光谱。他惊讶地发现，所有星云的光谱都在红移。换言之，这15个随机选取的螺旋星云，全都在飞离地球。

这是人类第一次看到宇宙膨胀的迹象。从这个意义上讲，斯里弗才是发现宇宙膨胀的第一人。但由于斯里弗所在的罗威尔天文台没有大口径的望远镜，他很快就陷入了止步不前的困境。

正所谓“工欲善其事，必先利其器”。要想取得最具革命性的天文学突破，还是要靠最大的天文望远镜。当时全世界最大的天文望远镜在哪里呢？答案是我们已经很熟悉的美国威尔逊山天文台。

这次活跃在威尔逊山天文台这个舞台上的，是我们的一个老熟人。他就是美国大天文学家哈勃。此前，哈勃已经利用标准烛光，发现银河系只是一个小小的宇宙孤岛，这让他成为天文学界的超级巨

星。

1928年，哈勃在欧洲开会期间，听到了用多普勒效应测量遥远星系速度的学术报告。他随即想到这样的问题：遥远星系的径向速度与它们到地球的距离之间，到底有什么关系？

回到威尔逊山天文台后，哈勃开始研究这个问题。测量星系距离一直是哈勃的拿手好戏；但是测量星系的径向速度，哈勃就不是很熟悉了。所以，他决定找一个助手，来帮忙分析星系的光谱变化。他找的这个助手，叫米尔顿·赫马森（Milton Humason）。



赫马森的早年经历异常坎坷。由于家境贫寒，他14岁就辍学了。为了谋生，赫马森打过各种各样的零工。1908—1910年，他的雇主是威尔逊山天文台。他的工作是赶着驴队，把建筑材料和生活物资送上威尔逊山，以支持天文台的建设。在此期间，他认识了一个天文台工程师的女儿，并和她结了婚。

1917年，在岳父的推荐下，赫马森成了威尔逊山天文台的一名看门人。此后每天晚上，他都会去找天文台的工作人员学习天文摄影技术；没过多久，他就成了天文台最好的观测助手。3年之后，只有一纸小学毕业证书的赫马森被任命为威尔逊山天文台的正式职员；而到了1922年，他又被破格提拔为助理天文学家。

但高等教育的缺乏，还是给赫马森的学术生涯蒙上了一层阴霾。由于基础不佳和命运不济，他曾两次与重大发现失之交臂。

第一次发生在1919年。当时，受一位天文学家的启发，赫马森开

始在一个特定的天区搜索太阳系的第9颗行星，并且拍摄了一大堆的照片。他对第9颗行星的搜索，最后以失败而告终。到了1930年，也就是冥王星被发现的那一年，赫马森的两个朋友重新检查了他之前拍摄的照片。结果发现，赫马森早在11年前就已经拍到了冥王星；但悲剧的是，他自己没认出来，所以就丢掉了“冥王星之父”的殊荣。

第二次发生在1920年。那年夏天，赫马森在仙女星云中发现了几个很异常的天体：其亮度会出现周期性的变化。这让他不禁怀疑，自己找到了仙女星云中的造父变星。这个发现，比哈勃在仙女星云中找到造父变星、进而确定仙女星云不在银河系内的历史性突破，要早上好几年。兴奋不已的赫马森，立刻标记了这些异常星在仙女星云中的位置，并把结果拿给了沙普利看。但不幸的是，坚信银河系是宇宙全部的沙普利对赫马森的发现根本不屑一顾。他先是盛气凌人地向赫马森解释为什么这些异常星不是造父变星，随后拿出手绢把所有数据抹掉。在学术界大权威面前，赫马森没敢坚持自己原来的想法。这样一来，他就与20世纪最大的天文发现之一擦肩而过。

在经历了两次失之交臂以后，赫马森终于等到了属于自己的机会。1928年，他开始与哈勃合作，研究星系的运动速度与它们到地球距离之间的关系。两人分工合作：赫马森基于多普勒效应，测量遥远星系的运动速度；哈勃则基于标准烛光，测量这些星系到地球的距离。

1929年，哈勃和赫马森已经测量了46个星系的速度和距离。结果显示，所有的星系都在远离地球。由于其中一大半的星系数据都存在很大的误差，哈勃只采用了那些他觉得有信心的数据。而基于这些星系观测数据，哈勃发表了一篇名为《河外星云距离与其径向速度的关系》的论文。

但这篇划时代的论文，并没有把赫马森列为作者。正因为如此，赫马森后来并没有获得自己应得的荣誉，而仅仅被视为“哈勃背后的男人”。

这篇论文的核心结论见图10。此图的横轴代表星系到地球的距离，其单位是百万秒差距（100万秒差距约等于326万光年）；而纵

轴代表星系的径向速度，其单位是千米/秒。图中的众多圆点代表哈勃和赫马森测量的那些星系。从图中可以看出，星系的径向速度与它到地球的距离正相关：星系离地球越远，它的退行速度（即远离地球的速度）就越大。

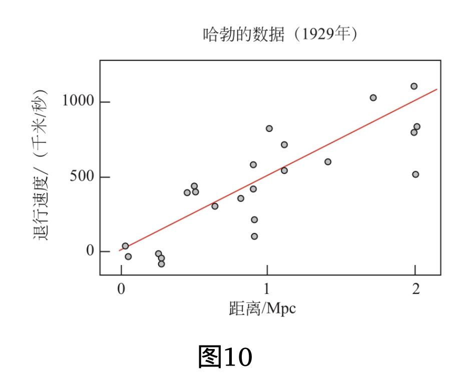



图10

但正相关仅仅是一个定性的结论。要是从定量的角度，确定此图中星系的退行速度与它们到地球距离之间的数学关系，就没那么容易了。此时的哈勃展现了他惊人的洞察力。他在图中画了一条穿过数据点的直线，然后宣称星系的退行速度正比于它们到地球的距离。

历史证明了哈勃的洞见。此后两年时间，他和赫马森一直在测量更遥远星系的速度和距离。他们找到的最遥远的星系，其退行速度高达20000千米/秒，而距离则超过1亿光年。1931年，哈勃与赫马森合写了一篇名为《河外星云的速度-距离关系》的论文。这篇论文的核心结论见图11。这回，星系的观测数据与哈勃画的直线完美契合。

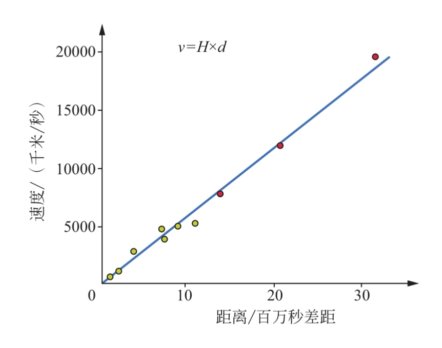


图11

星系的退行速度与它们到地球的距离成正比。这个结论，后来被称为哈勃定律。正是由于这条哈勃定律，人类终于意识到宇宙在膨胀。毫无疑问，这是天文学史上最伟大的发现之一。

哈勃定律到底意味着什么呢？答案是，它揭示出我们的宇宙必须满足宇宙学原理：宇宙在大尺度结构上是均匀且各向同性的。均匀是指，宇宙中的物质是均匀分布的；而各向同性是指，宇宙在各个方向上看起来都一样。这样一来，对于宇宙中任意位置的观测者，无论是什么时间，无论以什么角度，宇宙在大尺度结构上看起来都一样。

为了描绘这个宇宙图像，我们来做一个类比。想象有一个小圆球，突然发生了爆炸。这场爆炸把圆球炸成了许许多多大小一样的碎块，随即呈球形向外飞散。然后，你在一个飞散的碎块上，向位于球面上的其他碎块眺望（注意，你的视野始终局限在这个扩散的球面上，而无法望向其他的空间纬度）。这时你看到的碎块不断飞散、互相远离的画面，就满足哈勃定律和宇宙学原理。

在一个均匀且各向同性的宇宙中，所有的星系都在互相远离。这就是我们的宇宙正在放的电影。

现在，在脑海中把这部宇宙电影倒着放。你会发现所有的星系都在互相靠近。随着时间的不断推移，它们会变得越来越近，越来越近，最后恰好回到最初的一点。换句话说，在过去的某个时刻（现在一般认为是138亿年以前），宇宙中所有的物质都聚在一起，完全密不可分。你可以把这个最初的时刻定义为宇宙的起点。



这个图像，就是我们后面要重点介绍的宇宙大爆炸。

现在我们已经知道，哈勃定律揭示了宇宙会有一个起点，一个创生的时刻。那么，宇宙创生之初到底发生了什么？

欲知详情，请听下回分解。

---

(1)三棱镜是一种用玻璃制成、横截面为三角形的光学仪器。当光照到三棱镜的一个侧面之后，会先后发生两次折射，然后从另一个侧面射出。三棱镜对不同颜色的光会有不同的偏折程度：它对红光的偏折程度最小，而对紫光的偏折程度最大。

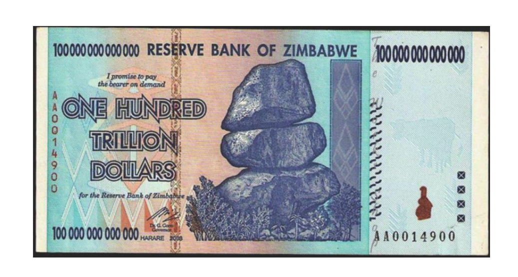
# 7 暴胀

上节课的结尾，我们提出了这样一个问题：宇宙创生之初到底发生了什么？

目前学术界最主流的答案是，暴胀。[1]

那什么是暴胀呢？且听我慢慢道来。

暴胀的英文是inflation，其本意是通货膨胀。通货膨胀说的是，在一段时间内，社会上流通的货币总量发生了显著的增长。而暴胀说的是，在宇宙创生后的一刹那，宇宙的体积发生了急剧地膨胀。




这个膨胀到底有多剧烈呢？答案是，在转瞬之间，宇宙总体积就膨胀了至少$1.6\times10^{60}$倍。这是什么概念呢？大概相当于一栋两层高的小楼，瞬间变得和整个银河系一样大。经过如此疯狂的膨胀之后，宇宙就变得和一个棒球差不多大小。然后，大爆炸才正式启动，最终让宇宙变成了我们今天看到的样子。

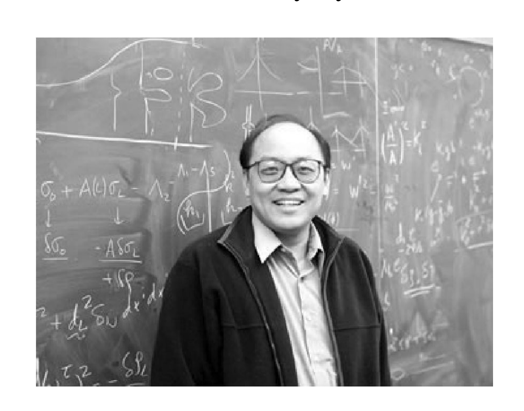
说到这里，你可能会觉得匪夷所思。为什么在宇宙创生之初要有一场暴胀？人类又是怎么发现此事的？

这就要从一个郁郁不得志的博士后的故事讲起了。他名叫艾伦·古斯（Alan Guth）。



1977年，多年研究粒子物理的古斯，跑到康奈尔大学物理系做第三期博士后。在此之前，他一直没做出足够好的科研成果，再加上运气不佳赶上了美国战后的婴儿潮，所以也一直没能找到大学助理教授的职位。如果再做不出好的科研成果，他就要被迫离开学术界了。

在康奈尔大学，古斯遇到了一个与他同病相怜的第三期博士后。此人是一个华人，名叫戴自海（Henry Tye）。



戴自海以前也研究粒子物理，后来转行做了宇宙学。他游说古斯，说粒子物理中最重要的课题已经被别人做得差不多了，不如和他一起做宇宙学。结果古斯根本没搭理他。戴自海也不气馁，还是隔三岔五就来游说古斯，这一游说就是两年。

直到1979年，事情才有了转机。那年年初，美国大物理学家斯蒂文·温伯格（Steven　Weinberg）到访康奈尔大学，并做了两场用粒子物理学理论研究宇宙学的演讲。温伯格是诺奖得主，同时也是美国粒

子物理学界的领袖。古斯一看连大名鼎鼎的温伯格都这么关心宇宙学，这才下定决心，与戴自海一起转战宇宙学。

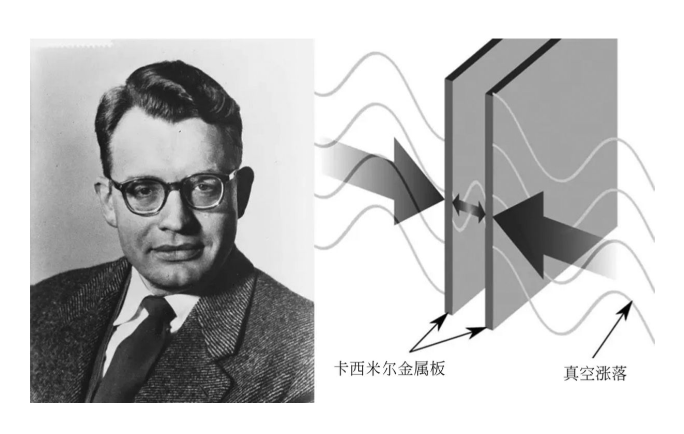
古斯和戴自海研究的，是一个与宇宙起源八竿子也打不着的课题，叫磁单极子问题。

简单解释一下什么是磁单极子。所有磁铁都有一个共同的特征：一定同时具有南北两极。即使把一块磁铁从中间一切两半，新得到的两块磁铁也会重新产生南极或北极。那么，有没有可能存在一种只有南极或北极的磁铁呢？理论上是可能的。像这种只有南极或北极的磁铁，就是磁单极子。

按照当时物理学界最流行的大统一理论，磁单极子在宇宙中应该是无处不在的，那为什么在真实世界中却连一个也找不到呢？这就是物理学界赫赫有名的磁单极子问题，同时也是古斯和戴自海决心挑战的课题。

他们的研究表明，解决磁单极子问题的关键是一个被称为“假真空”的概念。

为了解释什么是假真空，我们得先从真空的概念说起。很多人认为，真空就是一片什么东西都没有的空间区域。但我要告诉你的是，这种看法是错的。真空其实是有能量的。一个真空有能量的例子是著名的卡西米尔效应：由于真空有能量，处于真空中的两片不带电且相距很近的金属板之间会出现吸引力。而这个卡西米尔效应，已经得到了实验的证实。



一旦知道真空有能量，假真空就不难理解了。想象一座延绵起伏的大山，高的地方是山峰，矮的地方是山谷（如图12所示）。那么，如果在这座山上放一个小球，它在哪里可以保持静止呢？答案是山谷。现在把山的海拔高低视为空间本身的能量大小。凡是能让小球保持静止的山谷，全都处于真空的状态。换言之，真空就是能让置身其中的物体稳定存在的时空区域。很明显，真空也会有能量大小的差异，就像是山谷也会有海拔高低之分。其中能量最小的真空，对应于图中海拔最低的山谷，称为“真真空”；至于能量较大的真空，对应于图中海拔较高的山谷，则称为“假真空”。换句话说，假真空就是能量较高的真空。

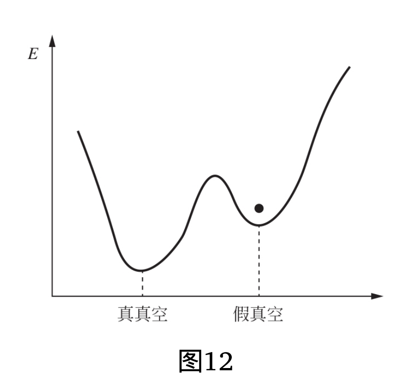



图12

戴自海率先意识到了一个最核心的问题：如果宇宙在诞生之初就处于一个假真空的环境里，它将会如何演化？这个问题，让人类迈出了破解宇宙创生之谜的关键一步。

但就在取得重大突破的前夕，戴自海却跑回中国参加了一个为期一个半月的学术会议。那是一个没有电子邮件和智能手机的年代。戴自海一回中国，就与古斯断了联系；等他重返美国的时候，古斯已经离开康奈尔去了斯坦福，而两人的合作也就此终止。

而正是在分开后的这段时期，古斯做出了宇宙学历史上最重大的突破之一。他发现，如果宇宙诞生在一个能量很高的假真空环境里，它就会被假真空的能量推动而向外膨胀。这就像是烤箱里的面团，会受到烤箱的热量而膨胀成面包。更关键的是，古斯发现在这种情况下，宇宙一定会发生指数式的急剧膨胀。而这正是前面说过的暴胀。

基于这个研究，古斯于1981年发表了一篇划时代的论文，正式提出了暴胀的概念。这篇论文，让古斯在学术界一夜暴得大名。遗憾的是，戴自海没被列为这篇论文的作者。

为什么古斯的暴胀理论能一夜爆红呢？原因在于，它一口气解决了三个困扰物理和天文学界的超级难题。

第一个是磁单极子问题：为什么我们完全找不到磁单极子？

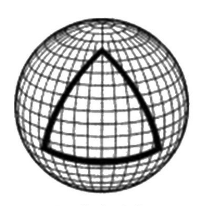
为了便于理解，我还是打个比方。将一把花瓣撒到一盆水中，你肯定能很轻易地在这盆水中把花瓣都找出来。如果把花瓣当成是磁单极子，磁单极子问题就是在问：为什么按理说很容易找到的花瓣，却完全找不到了？

暴胀理论对此问题的答案是，这盆水在转瞬之间就变得和太平洋一样大了。现在，你还能在太平洋中把这些花瓣都找出来吗？显然是做不到了。

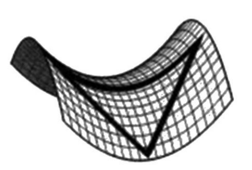
第二个是平坦性问题：为什么宇宙会如此平坦，以至于我们完全察觉不到空间本身的弯曲？

在日常生活中，我们对平坦和弯曲的概念一直局限在二维。比如说，桌子表面是平坦的，皮球表面是弯曲的，而两者都是二维的。这是因为，我们生活在三维空间，所以只能感知二维的平坦和弯曲。

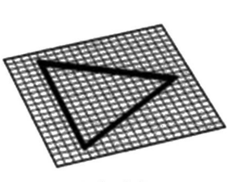
假设有一只生活在二维世界中的小蚂蚁，它该如何判断自己所处的二维空间到底是平坦还是弯曲呢？有个简单的办法：它可以在自己的二维空间内画一个三角形，然后测量此三角形的三个内角之和。如果内角之和等于180°，它所处的空间就是平坦的；如果内角之和大于180°，它所处的空间就弯成了一个球的形状；而如果内角之和小于180°，它所处的空间就弯成了一个马鞍的形状。



正曲率空间



负曲率空间



平坦空间

让我们回到三维世界。现在问题来了：怎么判断我们的三维空间到底是平坦的还是弯曲的呢？答案是，同样可以用画三角形的办法判断[(2)]。这意味着，类似于二维空间，三维空间也可以处于平坦或弯曲的状态。

现在你已经知道，从理论上讲，宇宙既可以是平坦的，也可以是弯曲的。因为平坦的状态只有一种，而弯曲的状态有无数种，所以从概率的角度来说，宇宙处于弯曲状态的可能性要大得多。但实际的天文观测表明，我们的宇宙是平坦的。这就很奇怪了。为什么宇宙会恰恰处于可能性最小的平坦状态呢？这就是所谓的平坦性问题。

暴胀理论对此问题的答案是，无论宇宙在创生之初是什么形状，暴胀都能把它抹平。举个现实生活中的例子。如果给你一颗小玻璃球，你一眼就能看出它是弯曲的。现在把这颗玻璃球变得和地球一样大，你还能一眼看出它并不平坦吗？显然就不行了（我们每天生活在地球上，根本察觉不到大地其实是球形的）。这意味着，半径越大的圆球，其弯曲程度就越小。暴胀迅速放大了整个宇宙的尺寸，从而把宇宙创生之初的空间弯曲给抹平了。

第三个是视界问题：为什么宇宙会这么均匀，以至于到处看起来都一样？

为了便于理解，我还是打个比方。有一群考生，在同一间教室里参加了一场两小时的考试。后来老师在批卷子的时候，发现所有人的答卷都完全相同，就连错误都一模一样。这该怎么解释呢？很明显，唯一的可能就是这些考生互相对了答案；或者说，他们彼此之间交流了信息。

现在有两批考生，其中一批人待在地球，而另一批人待在离地球4.3光年之遥的比邻星，他们都在同一时间参加了一场两小时的考试。你猜结果如何？所有人的答卷依然完全相同，就连错误都一模一样。

这就很诡异了。答卷一模一样，说明他们之间肯定交流了信息。但这两批考生相距4.3光年之遥，即使用速度最快的光，也要花整整4.3年才能把答案传过去。那他们是用什么办法，在短短两小时的时

间内就完成了信息的交流？

这个问题可以扩展到整个宇宙。目前的天文观测表明，在足够大的尺度上，宇宙中的物质分布地特别均匀，以至于到处看起来都一模一样。这说明，过去一定发生了信息的交流。但整个宇宙又这么大，即使是速度最快的光也不可能跑完，那它们到底如何完成信息的交流？换言之，宇宙如何完成超光速的信息交流？这就是所谓的视界问题。

暴胀理论对此问题的答案是，这两批考生一直待在同一间教室，只是由于暴胀让空间本身发生了急剧膨胀，从而让这两批考生之间的距离拉大到了4.3光年。其实在考试刚开始的时候，他们就已经完成了信息的交流。

由于一口气解决了磁单极子问题、平坦性问题和视界问题这三大科学疑难，暴胀理论一跃成为学术界最主流的宇宙创生理论。古斯也因此结束了颠沛流离的博士后生涯，被直接破格提拔为麻省理工学院的正教授。



但没过多久一些科学家就发现，古斯的理论其实存在着一个很大的缺陷。

前面讲过，在创生之初，宇宙处于一个能量很高的假真空环境。所以它会被假真空的能量推动做指数式的膨胀，这就是所谓的暴胀。不过，暴胀中的宇宙若想顺利变成我们今天看到的样子，必须要同时满足两个条件。

第一个条件，它要能及时地从假真空环境逃到真真空环境，从而

让暴胀结束。你可以把假真空当成是给宇宙吃的兴奋剂。运动员偶尔吃一点兴奋剂，能够大幅度提高运动成绩；但他要是每天都把兴奋剂当饭吃，肯定活不久。类似地，宇宙要是只在假真空环境下待一小段时间，就能靠暴胀解决一系列宇宙难题；但要是在假真空环境下待得太久，肯定会被暴胀扯得粉碎。

第二个条件，它要能在逃离假真空环境的过程中获得能量，从而为宇宙大爆炸（宇宙大爆炸的内容，我们下节课再详细介绍）点火。这是因为暴胀只是一个起点，之后还必须发生一次大爆炸，才能让宇宙逐渐变成今天的样子。要想引发大爆炸，就需要大量的能量。换句话说，要是不能在逃离假真空的过程中获得能量，宇宙大爆炸就发生不了。

知道了这两个条件，就可以讲讲古斯理论的缺陷了。按照古斯的理论，假真空是一个比真真空海拔更高的山谷。如果把刚刚诞生的宇宙当成一个皮球，那么它就诞生在一个假真空的山谷中。由于假真空山谷的周围都是比它能量更高的山坡，皮球很难从这个山谷中跑出去。

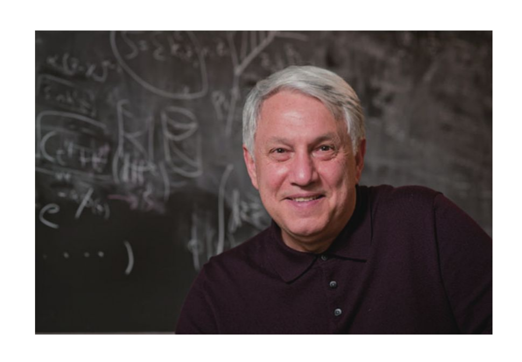
古斯认为，宇宙皮球可以用“量子隧穿”的方式逃离。为了便于理解，你可以把量子隧穿想象成《哈利·波特》中的一个咒语：“幻影移形。”只要挥舞魔杖并念出这个咒语，你就能从原先待的地方消失，并凭空出现在另一个地方。量子隧穿也能达到类似的效果。这样一来，原本诞生在假真空山谷中的宇宙皮球，在发生了量子隧穿以后，就能直接跑到真真空山谷了。

但问题在于，宇宙要是真的通过量子隧穿的方式逃离假真空，那它就无法获得任何能量了。做个类比，如果一个皮球从山顶上滚下来，那么当它滚到山脚时，会达到最大的速度。这是因为，皮球在山顶的重力势能会转化为它在山脚的动能。但如果皮球用幻影移形咒下山，那么它的运动状态就不会发生改变：下山前静止，下山后也还是静止。换句话说，皮球用量子隧穿的方式下山，就无法获得任何能量。

同样的道理，宇宙要是通过量子隧穿的方式，从假真空环境跑到

真真空环境，也无法获得任何能量。换句话说，按照古斯的理论，根本没办法为宇宙大爆炸点火。

最早化解这个危机的人，是暴胀学派的二号人物安德烈·林德（Andrei Linde）。而林德提出的解决之道，叫作慢滚暴胀理论。



慢滚暴胀理论说的是，宇宙并非诞生于一个假真空的山谷，而是诞生于一个假真空的山顶平台。很明显，皮球在山顶平台上也能保持静止，所以这个山顶平台同样可以被视为真空。



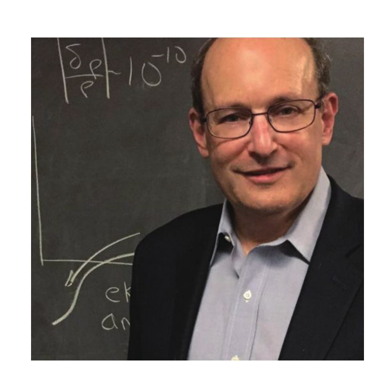
山顶平台的边缘，是通往真真空山谷的山坡。这样一来，皮球要离开山顶，就不再需要依靠量子隧穿，而可以沿着山坡正常地滚下来。换句话说，如果宇宙诞生之初就位于一个假真空的山顶平台上，那么它就能沿着平台边缘的山坡直接滚落到真真空的山谷，并自然而然地获得为大爆炸点火的能量。让宇宙从假真空平台上慢慢滚下来的暴胀，就是所谓的慢滚暴胀。

慢滚暴胀概念的提出，补齐了暴胀理论的最后一块短板，从而让

暴胀理论成为一个真正意义上的诺奖级的发现（时至今日，暴胀理论已经拿遍了除诺贝尔奖以外的所有科学界的大奖）。

最后，我再介绍暴胀理论的一个特别震撼的推论。

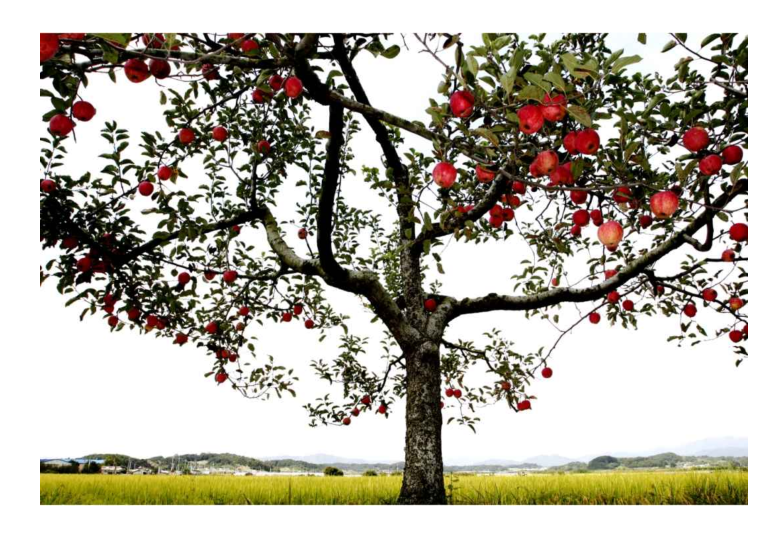
1983年，美国物理学家保罗·斯泰恩哈特（Paul Steinhardt）打开了一个暴胀理论的潘多拉魔盒，也就是所谓的永恒暴胀。它说的是，暴胀一旦开始，就永远都不会结束。



为了更好地解释永恒暴胀，我还是打一个比方。前面说过，要想制造一次暴胀，关键是要有一个假真空，也就是一个能量较高的真空。现在把假真空想象成一棵巨大无比的苹果树。苹果树的养分能在枝头结出苹果，就像假真空的能量能在某个空间区域造出一个暴胀的宇宙。当苹果长到足够大的时候，就会掉落；然后苹果树又可以结出新的苹果。类似地，当宇宙膨胀到足够大的时候，就会掉到真真空的山谷，然后假真空又可以制造新的暴胀宇宙。换句话说，永恒暴胀理论认为，假真空是一棵能够不断结出宇宙的苹果树，而我们的宇宙只是它结出的众多苹果中的一个。

要想理解永恒暴胀理论所带来的巨大冲击，不妨先回顾一下历史。400多年前，望远镜的发明让学术界意识到，我们的太阳并非银河系中唯一的恒星；100多年前，标准烛光的发现让学术界意识到，我们的银河系并非宇宙中唯一的星系；而1983年，永恒暴胀理论的提出则让学术界意识到，就连我们的宇宙也不见得是唯一的宇宙。这就是所谓的多元宇宙图像（按照弦论的观点，永恒暴胀所能创造的宇宙数量，大概是10的500次方的量级。说得更准确一点，总共有10的

500次方颗能产生宇宙的“苹果树”）。



时至今日，绝大多数的宇宙学家都已经接受了多元宇宙的概念。特别有趣的是，这个诡异的概念居然在公众间也颇受欢迎。比如说，美国著名的漫画公司漫威，就把它旗下诸多超级英雄所处的世界称为漫威多元宇宙。

我们已经介绍了宇宙创生后发生的第一件事，也就是暴胀。那么，暴胀之后宇宙又发生了什么呢？

欲知详情，请听下回分解。

---

(1)除了暴胀以外，理论家们还提出过火劫、反弹、弦气等宇宙起源理论。限于篇幅，这里就不介绍其他三种理论了。

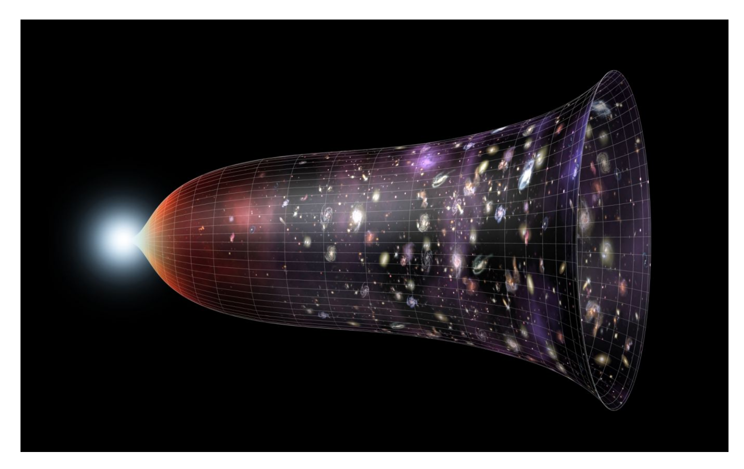
(2)“数学王子”高斯（Gauss）就曾这么干过。他是世界上最早怀疑我们生活的三维空间其实并不平坦的人之一，为此他还专门跑到德国的深山里测过三角形的内角和。不过此事，高斯是偷偷摸摸干的。因为他怕别人知道以后，会嘲笑他是神经病。

# 8　宇宙大爆炸

上节课的结尾，我们提出了这样一个问题：暴胀之后宇宙又发生了什么呢？

答案是，宇宙大爆炸。

人类发现宇宙大爆炸的历史，得从一位比利时的天主教神父说起。他叫乔治·勒梅特（George Lemaître）。





勒梅特参加过第一次世界大战，并因为作战英勇而获得过铁十字勋章。“一战”后，他上了一所神学院，并被任命为天主教牧师。随

后，他利用比利时政府提供的奖学金，先后前往剑桥大学、哈佛大学和麻省理工学院留学，并拿到了博士学位。

1925年，在比利时鲁汶大学找到固定教职的勒梅特，开始研究一个非常艰深的课题，那就是爱因斯坦（Einstein）的广义相对论。

广义相对论是爱因斯坦一生中最伟大的理论。它已经超越了牛顿万有引力定律，成了目前世界上最主流的引力理论[1]。

广义相对论最核心的公式是图13所示的爱因斯坦引力场方程。你不需要知道这个方程的细节。只要知道，方程的左式描述了宇宙的时空结构，而方程的右式描述了宇宙的物质分布；所以美国物理学家约翰·惠勒（John　Wheeler）认为广义相对论的本质是，“物质告诉时空如何弯曲，而时空告诉物质如何运动”。

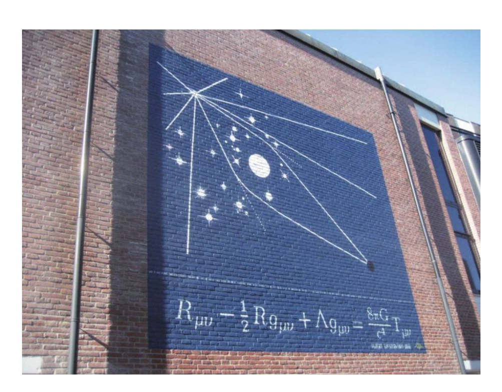



图13

重点来了。爱因斯坦最早写下这个引力场方程的时候，并不包括左式中的第三项。但他很快发现，在引力的作用下，宇宙将无法保持静止的状态。所以，爱因斯坦就在他的引力场方程中，引入了左式的第三项，也就是所谓的宇宙常数项。宇宙常数项能产生斥力，从而与引力达成平衡；这样一来，宇宙就可以处于永恒不变的静止状态。

勒梅特认为，宇宙常数项的引入非常突兀，根本就没什么道理。所以他想搞清楚，如果去掉这个宇宙常数项，会对宇宙学有什么影响。

勒梅特的研究表明，如果在爱因斯坦引力场方程中没有宇宙常数

项，那么宇宙就必须处于不断膨胀的状态；而且勒梅特预言，星系的退行速度应该与它们到地球的距离成正比。这恰好就是后来哈勃所发现的哈勃定律。

勒梅特并没有就此止步。他尝试倒放宇宙的电影。如果宇宙真的在膨胀，那么过去的宇宙一定比现在的宇宙要小。随着时间的倒流，宇宙会越来越小，直到把所有的天体都挤进一个超小型的宇宙。勒梅特就把这个最初的超小型宇宙称为“原始原子”。

一些大质量的原子（例如铀原子）会发生放射性衰变，从而分裂成较小的原子，并向外发射粒子和能量。所以勒梅特猜想，原始原子也会发生放射性衰变；衰变所放出的能量推动了宇宙的膨胀，而衰变所产生的物质凝聚成了星系和恒星。

宇宙起源于一个原始原子的放射性衰变，这就是勒梅特提出的“原始原子假说”。它正是宇宙大爆炸理论的雏形。

1927年，勒梅特在一次物理学会议上见到了爱因斯坦。他连忙凑到爱因斯坦身边，向这位科学巨人介绍自己提出的宇宙膨胀模型和原始原子假说。

结果，爱因斯坦完全不屑一顾。他告诉勒梅特，宇宙膨胀并不是什么新鲜事物；早在5年前，就已经有一个叫弗里德曼（Friedmann）的数学家提出了相同的理论（勒梅特此前并不知道弗里德曼的工作，所以宇宙膨胀的猜想是弗里德曼和他各自独立地提出的）。至于原始原子假说，爱因斯坦的评价是：“你的计算是正确的，但你的物理是可憎的。”

爱因斯坦的敌意和打压让勒梅特心灰意冷，而原始原子假说也被学术界打入了冷宫。但没过几年，勒梅特就“咸鱼翻身”了。这是因为，哈勃和赫马森发现的哈勃定律竟然与勒梅特的理论预言一模一样。这样一来，勒梅特就得到了包括英国大天文学家恶瑟·爱丁顿（Arthur Eddington）在内的一众学术界大佬的支持。最后，就连爱因斯坦都放弃了自己的静态宇宙模型，宣称引入宇宙常数项是他“一生中最大的错误”。[(2)]

但后来人们意识到，勒梅特的原始原子假说依然存在着一个很大

的缺陷：它根本无法解释宇宙中主要化学元素的丰度。

为了讲清楚这个问题，我们先从人们比较熟悉的化学元素周期表说起。这张表记录了人类目前发现的所有化学元素，其中排在前两位的元素，是氢和氦[(3)]。

天文观测表明，氢和氦的质量能占宇宙中所有化学元素总质量的99%；而氢和氦的质量之比，正好是3：1。为什么氢和氦的质量之比正好是3：1呢？这就是所谓的宇宙元素丰度问题。

最早破解这个超级难题的是一位传奇人物。他就是俄裔美籍物理学家乔治·伽莫夫（George Gamow）。



伽莫夫本科就读于列宁格勒大学，师从于我们前面提到过的、最早指出宇宙可能在膨胀的俄国数学家弗里德曼。不过，当时的伽莫夫对宇宙学毫不关心，他真正感兴趣的是量子力学及核物理。拿到博士学位以后，他跑到哥本哈根大学和剑桥大学做博士后，并在核物理的领域做出了世界级的成果。一家苏联的报纸对此进行了专题报道，并宣称："一位苏维埃学者向西方表明，在俄罗斯的土地上也能产生自己的柏拉图和牛顿。"

27岁那年，伽莫夫回到苏联，并成为了列宁格勒大学的教授。但没过多久，他就发现自己在苏联过得很不开心，所以就想带着自己的妻子离开苏联。

他曾和妻子一起前往一个位于苏联北部边境的小村庄，希望横渡北极水域跑到挪威，但因为有很多士兵把守边境线，不得不铩羽而

归。他也曾和妻子一起划一艘皮划艇，希望能横渡黑海跑到土耳其，结果遇到了一场大风暴，把他们的皮划艇打回了苏联的海岸。

后来在大物理学家玻尔（Bohr）和居里夫人（居里夫人一般指玛丽·居里，Marie Curie）的帮助下，伽莫夫利用一次出国开会的机会，成功地离开了苏联。30岁那年，他移民美国，成为华盛顿大学的教授。

当时的伽莫夫主要关心一个核物理领域的课题，即发生在恒星中心区域的氢核聚变（也就是4个氢核聚变为1个氦核的过程）。伽莫夫发现，恒星产生氦的速率非常慢：大概要花270亿年，才能让氢和氦的质量之比达到3：1。这意味着，恒星中心区域的氢核聚变，并不是宇宙中最主要的产生氦的方式。那么，宇宙中如此之多的氦，到底从何而来？

正是这个问题，把伽莫夫的目光引向了宇宙起源之谜。

伽莫夫猜想，宇宙创生之初的极端高温会把所有的物质结构都打碎。因此，充斥在极早期宇宙中的只能是一锅由质子、中子、电子和光子混合而成的“热汤”。伽莫夫把这锅热汤称为“ylem”。“ylem”是一个已被废弃的古英语单词，它的意思是“构成元素的原始物质”。

随着宇宙的不断膨胀，这锅由质子、中子、电子和光子混合而成的“ylem”的温度，也会不断降低。当宇宙温度降到某个临界值的时候，“ylem”就会开启氢核聚变过程；而当宇宙温度继续降到另一个临界值的时候，“ylem”就会终止氢核聚变过程。在此期间，宇宙就可以产生大量的氦。这个过程，就是所谓的“原初核合成”。

顺便多说一句。按伽莫夫的原意，这个原初核合成的过程其实就是宇宙大爆炸。

这是一个非常天才的构想。但问题是，计算原初核合成过程中发生的各种核反应，是一件极端复杂的事情。伽莫夫的数学不好，根本无力完成这么复杂的计算，所以他面临的是一种几乎绝望的困境。

直到1945年，伽莫夫才看到了走出这个困境的曙光。他遇到了一个堪称数学天才的年轻人，名叫拉尔夫·阿尔菲（Ralph Alpher）。

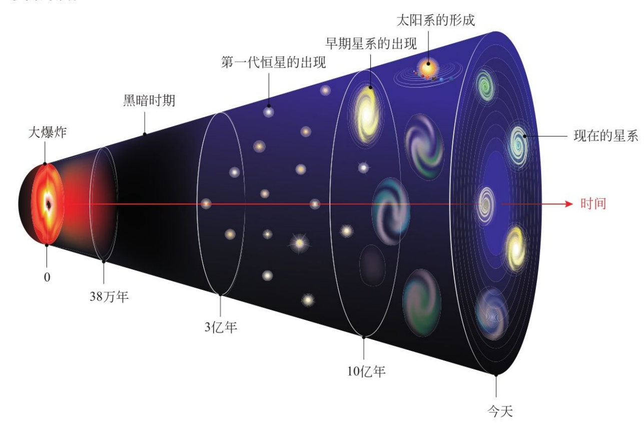
16岁那年，阿尔菲拿到了麻省理工学院（Massachusetts Institute of Technology, MIT）的全额奖学金。但不幸的是，在与MIT校友聊天的时候，阿尔菲不慎暴露了自己的犹太血统，这导致他的奖学金被直接取消。无奈之下，阿尔菲只好选择白天工作，晚上念华盛顿大学的夜校。最终，他通过这样的方式，拿到了自己的学士学位。

正是在此期间，伽莫夫遇到了阿尔菲。这个年轻人的数学才华，让伽莫夫眼前一亮。因此，他立刻将阿尔菲招收为自己的博士生。

伽莫夫和阿尔菲对原初核合成的研究持续了整整3年。他们完成了一个跨学科的壮举：用核物理的知识来研究宇宙起源。最终的计算结果表明，在原初核合成的末期，差不多每10个氢原子核能生成1个氦原子核。这样一来，当原初核合成结束后，氢和氦的质量之比就会达到3：1。这意味着，宇宙大爆炸理论能够完美地解释氢元素和氦元素的丰度。这是继成功预言哈勃定律以后，宇宙大爆炸所取得的又一次重大胜利。

为了宣布这个重大突破，伽莫夫和阿尔菲用他们最终的计算结果和结论，写了一篇名为《化学元素的起源》的论文。这篇论文在1948年4月1日，也就是愚人节的那天，发表在了《物理评论》杂志上。这是一篇很有愚人节特色的论文。因为伽莫夫把自己一个与此论文毫无关系的朋友［此人是汉斯·贝特（Hans Bethe），1967年诺贝尔物理学奖得主］，强行塞进了作者的列表。伽莫夫之所以要这么做，是为了让此文三个作者的名字——阿尔菲、贝特、伽莫夫，连起来能凑成αβγ。因此，后人也把这篇愚人节论文，称为αβγ论文。

这篇αβγ论文，无疑是宇宙学史上的一座丰碑。它证明了一锅由质子、中子、电子、光子混合而成的“热汤”，就足以最终演变成我们今天看到的宇宙。




基于这篇论文，阿尔菲开始申请他的博士学位。最后的博士答辩，吸引了一大批华盛顿的记者。他们注意到了阿尔菲在答辩时说的一个结论：氢和氦的原初核合成，只发生在最初的300秒以内。此后数天，这句话成了多家美国报纸的头条新闻。《华盛顿邮报》就写道："世界始于最初的5分钟。"

后来的研究表明，宇宙创生的时间（即原初核合成结束的时间）大概是3分钟。接下来，我就用现代的观点，为你播放一下这部宇宙创生的电影。

在138亿年前的某个时刻，宇宙诞生。此时，宇宙的体积为0，温度无限高，密度无穷大，这就是宇宙奇点。目前，人类对宇宙奇点还一无所知。

在诞生后的$10^{-43}$～$10^{-35}$秒里，宇宙处于普朗克时期。在此期间，自然界中的四种基本力，即引力、电磁力、强核力和弱核力，还属于同一种力，即超力（superforce）。[(4)]到了$10^{-35}$秒，宇宙温度下降到$10^{27}$℃，此时发生第一次宇宙相变[(5)]，让引力从超力中分离出来。

在诞生后的$10^{-35}$～$10^{-32}$秒里，宇宙处于暴胀时期。在此期间，宇宙总体积至少膨胀了$1.6\times10^{60}$倍，相当于从一栋两层小楼瞬间变得和整个银河系一样大。上一节讲过，由于这个急剧地膨胀，磁单极子问题、平坦性问题和视界问题这三大疑难，全都迎刃而解。这段时期的另一件大事，是发生了第二次宇宙相变，让强核力也从超力中分离出

来。到了$10^{-32}$秒，宇宙脱离了假真空的环境，暴胀也随之终止。

在诞生后的$10^{-32}$～$10^{-10}$秒里，宇宙处于物质形成时期。此前的暴胀让宇宙温度急剧下降。但是宇宙在脱离假真空环境的过程中又获得了大量的能量（参阅上一节讲的慢滚暴胀理论），而这些能量又为宇宙重新加热。此后，宇宙中充斥着正反物质[(6)]，主要是夸克（夸克是一种比质子和中子更基本的微观粒子。事实上，质子和中子都是由三个夸克构成的）、反夸克、电子和反电子。随着温度的下降，正反物质会发生湮灭（一个正物质粒子与它的反物质粒子相撞后，会一起消失，并发出两个光子。这个现象就是湮灭）。由于某种原因，在宇宙中正物质粒子的数量比反物质粒子的数量要多10亿分之一。等正反物质互相湮灭后，这多出来的10亿分之一的物质，就逐渐演化成了我们今天看到的宇宙。到了$10^{-10}$秒，宇宙温度下降到$10^{15}$℃，此时发生了弱电相变，让弱核力和电磁力也分离开来。

在诞生后的$10^{-10}$～1秒里，宇宙处于夸克禁闭时期。在此期间，夸克互相结合，产生质子和中子。这就是伽莫夫设想的那锅由质子、中子、电子、光子混合而成，名为“ylem”的热汤的起源。到了秒的时候，宇宙温度下降到$10^{10}$℃，随之启动氢核合成过程。

而在诞生后的1秒至3分钟里，宇宙处于原初核合成时期（按照αβγ论文的原意，原初核合成就等同于宇宙大爆炸）。在此期间，由质子、中子、电子、光子混合而成的热汤一直在进行核聚变反应。到了3分钟的时候，核聚变终止。此时宇宙变成了一个由氢和氦构成的火球，且氢和氦的质量之比为3：1；此外，在火球中也有少量的锂元素。

时至今日，这个看起来玄乎其玄的宇宙大爆炸理论，早已成为学术界最有名、也最成功的宇宙起源理论。

为什么宇宙大爆炸理论能取得如此巨大的成功？

欲知详情，请听下回分解。



---

(1)由于篇幅所限，这里就不展开介绍广义相对论了。对广义相对论物理图像感兴趣的读者，可以参阅我之前写的《宇宙奥德赛：漫步太阳系》一书的4.2节。

(2)诡异的是，到了20世纪末，情况竟然再次发生反转。以今天的眼光来看，宇宙常数不但不是爱因斯坦犯的错误，反而有可能是他最伟大的洞见。

(3)氢原子由一个氢原子核和一个电子构成，氢原子核包含一个质子；氦原子由一个氦原子核和两个电子构成，氦原子核包含两个质子和两个中子。

(4)很多物理学家相信，引力、电磁力、强核力和弱核力在宇宙创生之初是统一的；随着宇宙温度的下降，这四种力就逐一分离出来。

(5)相变是指在某种临界条件下，事物从一种状态突变到另一种状态的现象，例如，温度降到0℃时的水变冰。

(6)反物质与物质的唯一区别，是它们所带电荷的符号不同。比如说，质子带正电荷，而反质子带等量的负电荷；电子带负电荷，而反电子带等量的正电荷。

# 9 宇宙微波背景辐射

上节课的结尾，我们提出了这样一个问题：为什么宇宙大爆炸理论能取得如此巨大的成功？

原因是，人类后来找到了宇宙大爆炸最核心的理论预言，即宇宙微波背景辐射。

人类探索宇宙微波背景的故事，得从一个之前已经登过场的人物讲起。他就是αβγ论文中的那个α，拉尔夫·阿尔菲。



前面我们讲过，伽莫夫为了把作者列表凑成αβγ，强行把他的朋友贝特拉进了作者名单。伽莫夫开的这个玩笑，让一个人感到极端不满，此人就是阿尔菲。

阿尔菲担心，一旦加入贝特的名字，就会大大降低学术界对自己的评价。大家会认为，这篇提出宇宙大爆炸和原初核合成概念的论文主要是伽莫夫和贝特的功劳，自己只是在给这两个大科学家打工罢了。

所以，阿尔菲就需要甩开自己的导师伽莫夫，单独完成一项关于宇宙大爆炸的研究工作。这样一来，他才能证明自己是一个足以独当一面的科研人才。

阿尔菲开始与一位叫罗伯特·赫尔曼（Robert　Hermann）的同事合作，进一步研究宇宙大爆炸理论。他们最关心的问题是，宇宙大爆炸有没有一个能被天文观测检验的理论预言（氢和氦的宇宙丰度不算。因为早在宇宙大爆炸理论提出前，人们就已经知道了宇宙中氢和氦的质量之比是3：1）。换句话说，他们想知道，有没有能被天文观测看到的宇宙大爆炸的遗迹。

说到这里，你可能会觉得匪夷所思了。一场发生在138亿年前的大爆炸，怎么可能会留下今天还能看到的遗迹？但是阿尔菲和赫尔曼的研究表明，还真有一样东西能留得下来，那就是宇宙大爆炸的火球所发出的光。

不过，这并不是宇宙创生之初所发出的光。这是因为，创生之初的火球过于炽热。在如此高温下，原子核无法与电子结合形成原子。此时的宇宙，就是一锅由原子核和自由电子混合而成的等离子体汤（等离子体是不同于固态、液态、气态的第四种物态。其核心特征是，原子核和电子各自独立，无法结合成原子）。

在这个充斥着等离子体汤的宇宙火球中还有大量的光子。光子非常容易与带电粒子发生相互作用。这意味着，只要宇宙的温度较高，就会让光处于一种被囚禁的状态。

一直到宇宙诞生后38万年后，宇宙火球的温度才下降到3000℃。此时，原子核才能与电子结合形成原子，这就是所谓的宇宙复合时期。此后，就没有带电粒子来干扰光子的运动了。这样一来，光就可以在宇宙中自由传播了。

宇宙诞生38万年后发出的光，在经历了138亿年的悠悠岁月后，终于到达地球。这些光的波长，全都被宇宙膨胀拉伸到微波的波段。这些在宇宙诞生38万年后发出的、目前已处于微波波段的光，就是所谓的宇宙微波背景。而宇宙微波背景，就是能被天文观测看到的、宇宙大爆炸的理论预言。

1948年年末，阿尔菲和赫尔曼在《自然》杂志上发表了一篇论文。这篇论文首次提出宇宙微波背景的概念。他们指出，如果宇宙大爆炸理论是对的，那么我们就能在地球上接收到来自宇宙各个方向、

波长为毫米量级的微波信号。

从某种意义上讲，这是一篇比αβγ论文还要重要的史诗级论文。它让看似虚无缥缈的宇宙大爆炸理论登堂入室，成为一门真正意义上的现代科学。

不过，作为宇宙大爆炸理论的先驱，伽莫夫、阿尔菲和赫尔曼的前路依然遍布荆棘。

有个流传甚广的说法：如果你领先一个行业1年，就会成为这个行业的先驱；如果你领先一个行业10年，就会成为这个行业的先烈。

伽莫夫、阿尔菲和赫尔曼，就成为了宇宙学的先烈。

在此后的5年时间里，他们一直尝试说服天文学家去寻找宇宙微波背景。但是，根本没有任何人回应。由于天文学界对宇宙大爆炸理论的巨大冷漠，他们在1953年放弃了对宇宙起源的进一步研究。伽莫夫还是留在学术界，但把大量精力都拿去写科普书了。至于阿尔菲和赫尔曼，则先后放弃了学术界的生涯，转行去了工业界。

随着伽莫夫、阿尔菲和赫尔曼的离去，宇宙大爆炸理论就陷入了漫长的沉寂。直到11年后，由于另一个人的出现，宇宙大爆炸理论才得以重见天日。此人就是美国天文学家罗伯特·迪克（Robert Dicke）。

迪克是普林斯顿大学的天文系教授，同时也是一个拥有50多项专利的发明家。他发明过一种叫迪克辐射计的装置，该装置能以很高的灵敏度探测波长为1厘米的微波信号。而这个迪克辐射计，后来也成了所有射电望远镜的核心设备。



到了20世纪60年代，迪克突然对宇宙起源问题产生了兴趣。他和自己的得意弟子詹姆斯·皮布尔斯（James Peebles）一起，研究了宇宙创生之初产生的火球。而他们最关心的课题是，这个火球可不可能留下足以被射电望远镜看到的遗迹。



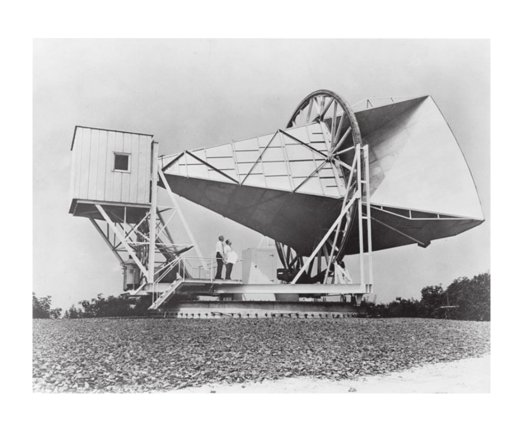
迪克和皮布尔斯的研究，最后变成了一篇两人合写、于1964年发表的论文。在这篇论文中迪克和皮布尔斯预言，宇宙大爆炸一定会留下宇宙微波背景，而后者完全可以被射电望远镜看到。不幸的是：在写这篇论文以前，迪克和皮布尔斯并没有做充分的文献调研；所以，他们没有意识到伽莫夫、阿尔菲和赫尔曼早已做过相同的研究。因此，这篇论文就对伽莫夫、阿尔菲和赫尔曼的工作只字未提。

迪克并不打算止步于此。他决定自己造一个射电望远镜，来搜寻宇宙大爆炸留下的宇宙微波背景。如果真能找到这个宇宙微波背景，那将成为天文学史上的一座丰碑。但迪克的梦想，却因为他在一年后接到的一通电话，而化为了泡影。

打这通电话的是贝尔实验室的两个研究人员，阿诺·彭齐亚斯（Arno Penzias）和罗伯特·威尔逊（Robert Wilson）。



20世纪60年代初，贝尔实验室在新泽西州的克劳福德山上，造了一个跨度6米的“大喇叭”。这是一个可以360°旋转的射电天线，最初被设计用来接收一颗军事卫星发回地球的无线电波信号。项目结束后，这个射电天线被改造成了一个射电望远镜。彭齐亚斯和威尔逊的工作，就是用这个喇叭状的射电望远镜扫描天空，进而研究天上的各种射电源（即能发出无线电波的天体）。



万万没想到，这成了两人“噩梦”的开始。

从建好望远镜的第一天起，彭齐亚斯和威尔逊就遇到了一个巨大的麻烦：他们的“大喇叭”会持续不断地收到一种特定频率的噪声，类似于电视机收不到电视台信号时出现的那种雪花屏。更诡异的是，不管他们如何调整“大喇叭”的方向，这种神秘的噪声都不会消失。换句话说，这种神秘噪声来自于宇宙的各个方向，而且完全不受昼夜和季节因素的影响。

一种神秘的无线电波信号，竟然持续不断地从宇宙的各个方向传来。这让彭齐亚斯和威尔逊不禁怀疑，是这个射电望远镜本身出了问题。他们花了一年的时间，检查了射电望远镜的每一个环节，包括所有的器件、线路和接口。经过仔细的排查，他们终于找到了“症结”所在：一对鸽子在大喇叭里筑了窝，并在射电天线上拉了很多白色的鸽子屎。彭齐亚斯和威尔逊猜测，正是这些落在天线上的鸽子屎，导致了那种神秘的噪声。

所以，彭齐亚斯和威尔逊就抓住了这对鸽子，然后把它们带到了

50千米外的地方放飞。但问题是，鸽子有归巢的本能，没过多久这对鸽子又飞回了大喇叭。无奈之下，两人只好找了个猎人，以粗暴的方法解决了鸽子的问题。

然后，彭齐亚斯和威尔逊就清理了鸽子窝，并对这个大喇叭做了一番大扫除。经过一年的检查、清洁和重新布线，彭齐亚斯和威尔逊又打开了他们的射电望远镜。结果他们目瞪口呆地发现，那个困扰了他们整整一年的神秘噪声，依然存在。

就在彭齐亚斯即将崩溃之际，一个学术圈的朋友给他带来了福音。这个朋友给彭齐亚斯寄去了一篇论文。它正是迪克与皮布尔斯合写的那篇预言宇宙微波背景辐射的论文。

看完这篇论文之后，彭齐亚斯顿时醍醐灌顶。他终于明白，已经折磨他整整一年的神秘噪声，并不是命运的诅咒，而是上天的眷顾。

于是，彭齐亚斯就和威尔逊一起，给当时正和学生吃午饭的迪克打了一通电话。他们告诉迪克，他们已经发现了迪克想要寻找的东西。放下电话后，迪克神情落寞地告诉自己的学生："我们被别人抢先了。"

1965年夏天，彭齐亚斯和威尔逊在《天体物理学期刊》上发表了一篇划时代的论文。在这篇论文中，彭齐亚斯和威尔逊宣布，他们发现了一种来自宇宙各个方向的神秘微波噪声。文章花了不少篇幅来描述他们是如何排查仪器故障的。为了尽量文雅一点，凡是提到白色鸽子屎的地方，全都改成了"白色介电材料"。

与此同时，迪克的团队也在同一家杂志上，发表了姐妹篇论文：他们明确地指出，彭齐亚斯和威尔逊发现的神秘微波噪声，就是宇宙大爆炸的遗迹，即宇宙微波背景。对宇宙大爆炸学派来说，这无疑是一个最辉煌的胜利。

彭齐亚斯和威尔逊的重大发现，也激起了公众的强烈兴趣。就连《纽约时报》都以头版头条报道了宇宙微波背景的发现。报道引述了彭齐亚斯本人对此的描述："当你今晚走到户外，并摘下帽子，你的头皮就能感受到大爆炸带来的一丝温暖。如果你有一个品质良好的调频收音机，而且你站在两个微波中继站之间，你就会听到'嘶-嘶-嘶'的

声音。你可能听到过这样的哔哔声。它像是一种抚慰，有时又像是海浪的拍击声。你听到的声音，大约有千分之五来自于数十亿年前传来的宇宙噪声。”

宇宙微波背景的发现，让沉寂多年的宇宙大爆炸理论在一夜之间就登上了科学的神坛。听到这个消息后，伽莫夫、阿尔菲和赫尔曼也回来了，但他们的喜悦中却夹杂着苦涩。因为他们早年的开创性贡献已经被世人遗忘了。就连发表在《天体物理学期刊》上的那两篇论文，也对他们3人的贡献只字不提。

伽莫夫试图利用一切机会在公共场合发声，以确立自己团队在宇宙大爆炸和宇宙微波背景领域的优先权。举个例子，有人在一个学术会议上问伽莫夫：彭齐亚斯和威尔逊发现的宇宙微波背景，是否确实是他、阿尔菲和赫尔曼曾经预言过的现象。伽莫夫傲然地回答道：“好吧，让我打个比方。我在这附近掉了一枚硬币。现在有人在我掉硬币的地方找到了一枚硬币。我知道所有的硬币看起来都一样，但我相信这枚硬币就是我掉的那枚。”

后来彭齐亚斯也知道了此事。他给伽莫夫写了封信，要求伽莫夫提供能证明自己优先权的更多信息。伽莫夫在回信中，详细地介绍了自己团队之前所做的一系列研究工作。在信的结尾，伽莫夫不无讽刺地写道：“所以你看，世界并非始于万能的迪克。”

而阿尔菲的反应就更激烈了。他曾经向一个记者公开表达自己对彭齐亚斯和威尔逊的愤慨：“我能不失望吗？他们考虑过我的感受吗？他们甚至从未邀请我们去看看那个该死的望远镜！”

虽然迪克和皮布尔斯后来承认了伽莫夫、阿尔菲和赫尔曼的贡献，但是伤害已经造成。这场争吵的结果，让几乎所有人都成了输家。

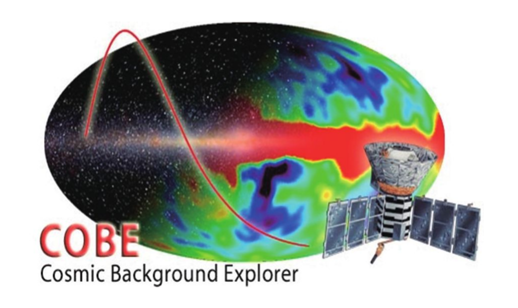
宇宙微波背景的发现，让彭齐亚斯和威尔逊获得了1978年的诺贝尔物理学奖。但是对大爆炸宇宙学的建立做出了更大贡献的那5位理论家（即伽莫夫、阿尔菲、赫尔曼、迪克和皮布尔斯），却因为种种争议，迟迟没能戴上诺贝尔奖的桂冠。

直到2019年10月8日，诺贝尔物理学奖才被授予给5人中最年轻

的皮布尔斯。这回，终于不再有任何争议了，因为其他4人都已经不在人世了。

我们已经讲完了人类发现宇宙大爆炸的漫长而曲折的历史。最后，我再介绍一项关于宇宙微波背景的研究。

1989年，由美国科学家约翰·马瑟（John Mather）和乔治·斯穆特（George Smoot）领导的一个科研团队，发射了一颗名为“宇宙背景探测者”（cosmic background explorer, COBE）的卫星。发射这颗卫星的目的，是要以更高的精度来探测宇宙微波背景。




1992年，COBE团队宣布他们有了一个重大发现。这个发现后来让马瑟和斯穆特拿到了2006年的诺贝尔物理学奖。

图14就是让马瑟和斯穆特拿到诺奖的重大发现。

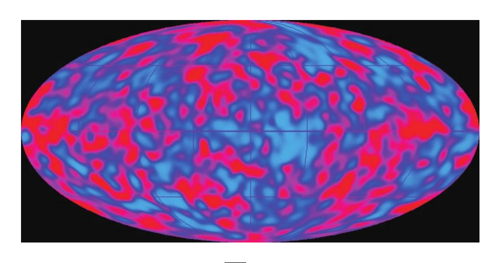


图14

你可以把它想象成一张地图。准确地说，这是宇宙诞生38万年后的宇宙地图。图中红色的部分，表示宇宙中物质密集的区域；而蓝色的部分，表示宇宙中物质稀疏的区域。也就是说，利用COBE卫星，马瑟和斯穆特等发现，诞生38万年后的宇宙存在着物质分布的微小不均匀性。这种物质分布的微小不均匀性，就是所谓的宇宙密度涨落。

那么，宇宙密度涨落导致的不均匀性到底有多微小呢？答案是，图中红色区域的物质密度比蓝色区域的物质密度要大10万分之一。

正是这区区10万分之一的密度涨落，最终演变成了我们今天所看到的恒星世界。

那么问题来了：宇宙密度涨落为什么会演变成恒星呢？

欲知详情，请听下回分解。

# 10　恒星的一生

上节课的结尾，我们提出了这样一个问题：宇宙密度涨落为什么会演变成恒星呢？

答案是，由于宇宙版本的马太效应。

马太效应源于《新约·马太福音》中的一则寓言：一个要出远门的人，把自己的财产托付给了三个仆人。第一个仆人分到了5000银元，第二个仆人分到了2000银元，而第三个仆人分到了1000银元。许久之后，主人回来了，和三个仆人算账。第一个仆人说："主人，我把您给的5000银元拿去做生意，又赚了5000银元。"主人听后很高兴，就给他升了职。第二个仆人说："主人，我把您给的2000银元拿去做生意，又赚了2000银元。"主人听后很满意，也给他升了职。而第三个仆人说："主人，我把您给的1000银元埋在了地里，现在它们全在这里。"主人听后勃然大怒，立刻收回了他的1000银元，然后全部交给了现在已有10000银元的第一个仆人。

后来受这则寓言的启发，美国学者罗伯特·默顿（Robert Merton）提出了马太效应的概念。马太效应说的是，人类社会存在赢家通吃的现象。通俗地说，富人会越来越富，而穷人会越来越穷。

诡异的是，马太效应不光适用于人类社会，还适用于整个宇宙：宇宙中物质比较密集的区域，会依靠自身较大的引力把周围的物质都吸过来；而吸过来的物质多了，又能够让引力变得更大。如此一来，就会形成良性循环，让该区域变得越来越密集，最终造成引力塌缩。(1)

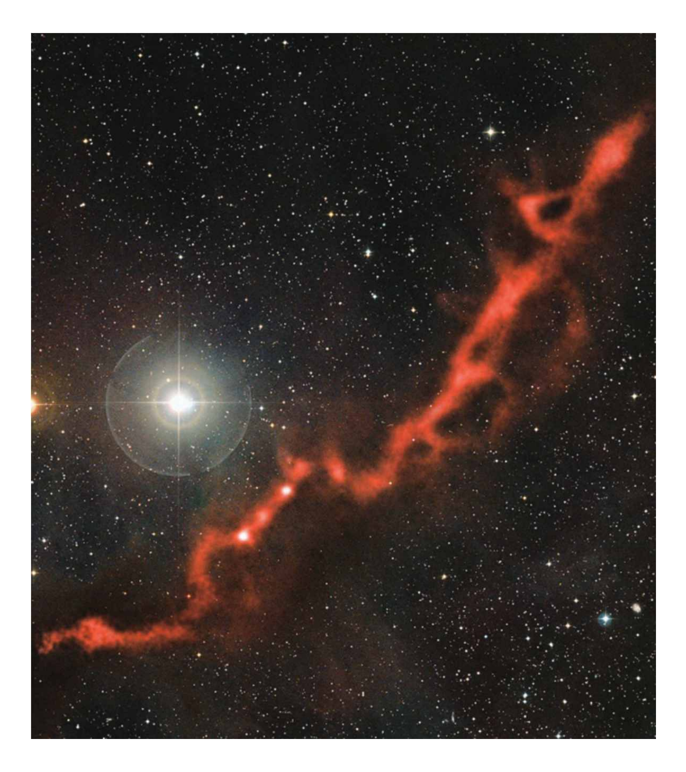
知道了宇宙版本的马太效应，我们就可以介绍恒星到底如何诞生了。恒星诞生的过程，可以分为三个阶段。

恒星诞生的第一阶段，是从宇宙密度涨落中产生分子云。

上节课已经讲过，在宇宙创生38万年后，出现了宇宙密度涨落：那些物质密集区域的密度，比物质稀疏区域的密度大10万分之一。这

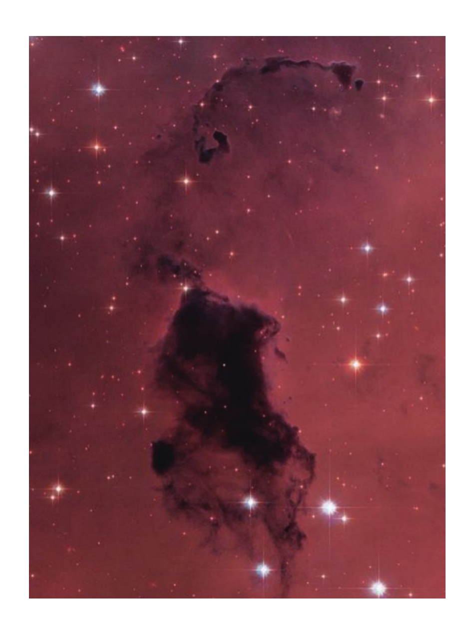
个初始的宇宙密度涨落就像一颗种子，在马太效应的滋养下会越长越大。换言之，由于马太效应，物质密集区域和物质稀疏区域之间的密度差异会越变越大。最后，就会制造出一片物质密度远大于宇宙平均密度的区域。这片大密度区域中的物质（主要是氢和氦）一般以分子的形式存在。所以，它就被称为分子云。

为了便于理解，你不妨把分子云当成是恒星的育婴室。而分子云还可以再细分成三类。



最大的分子云叫巨分子云，一般分布在几百光年的空间范围内，其质量约为太阳质量的几百上千万倍。下图就是一个典型的例子——金牛座巨分子云。

最小的分子云叫博克球状体，一般分布在不超过一光年的空间范围内，其质量约为太阳质量的几倍。下图就是一个位于NGC 281星云中的博克球状体。



介于巨分子云和博克球状体之间的是中等质量分子云，一般分布在几十光年的空间范围内，其质量约为太阳质量的几十倍甚至上百倍。下图就展示了一个最有名的例子——位于老鹰星云中的创生之柱。

恒星诞生的第二阶段，是从分子云中产生原恒星。

1947年，荷兰天文学家巴特·博克（Bart Bok）提出了一个假说：分子云会发生碎裂，从而形成一些分子云的碎块。每个碎块的中心都会出现一个非常致密的核心，而这个核心又会进一步吸引外围的物质。因此，分子云核心会被它的外围物质包裹起来，就像是一只被蚕茧包裹起来的蚕宝宝。

最初，分子云核心的温度相当低，大概只有10开尔文（相当于零下263℃）。因为分子云核心一旦升温，就会发出大量的电磁波；电磁波可以从外围“蚕茧”的缝隙中逃逸，从而把能量带走，让分子云核心的温度迟迟无法升高。



在温度很低的情况下，分子云核心向外扩张的压力远远小于其自身的引力。所以，分子云核心会处于加速收缩的状态。

随着外层物质越聚越多，外围的“蚕茧”会不断变厚。等“蚕茧”厚到能把电磁波全部拦截下来的时候，分子云核心的温度就可以显著上升了。当核心温度达到3000开尔文的时候，向外扩张的压力就能与引力达到平衡了。

这是一个关键的时点。此后，分子云核心的温度会进一步升高，让自己进入减速收缩的状态。这种处于减速收缩状态的分子云核心，就是所谓的原恒星。

为了便于理解，你不妨把原恒星当成是胚胎状态的恒星。图15就展示了一颗被称为“HOPS 383”的原恒星。

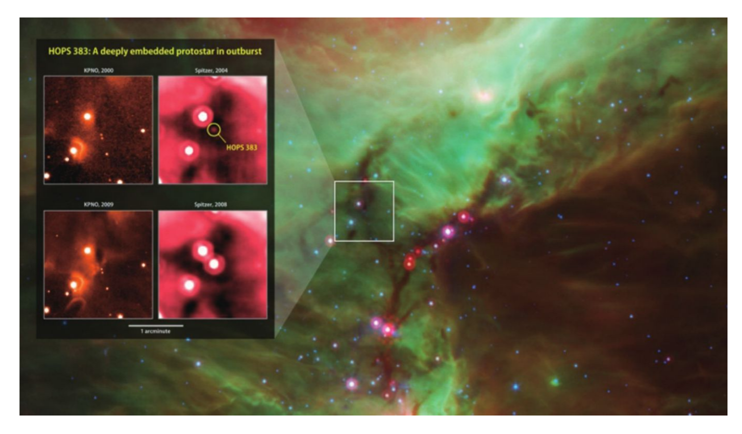



图15

恒星诞生的第三阶段，是从原恒星变成真正的恒星。

在这个过程中，会同时发生两件大事。

第一件大事，原恒星会继续地从包裹它的“蚕茧”中吸收物质。由于“蚕茧”中的物质是有限的，原恒星最后能吃掉整个“蚕茧”。

第二件大事，原恒星的温度会随体积的收缩而不断升高。当温度突破某个临界值的时候，就可以在原恒星的中心点燃氢核聚变。一旦氢核聚变被点燃，原恒星就会变成一颗真正的恒星。

完成这两件大事的先后顺序，决定了原恒星变成真正恒星的两种路径。

如果分子云碎块的体积比较小，就会形成一个小质量的原恒星，以及比较薄的外层“蚕茧”。在这种情况下，当原恒星把整个“蚕茧”都吃掉后，其中心依然没能点燃氢核聚变。此后，这个已经没有“蚕茧”包裹的原恒星会继续收缩，最终突破临界温度并点燃氢核聚变。这种路径，会形成一颗小质量恒星。

如果分子云碎块的体积比较大，就会形成一个大质量的原恒星，以及比较厚的外层“蚕茧”。在这种情况下，原恒星还没来得及把外层“蚕茧”吃掉，其中心就已经点燃了氢核聚变。氢核聚变释放的巨大能量，会把外围“蚕茧”直接吹散。这种路径，会形成一颗大质量恒星。

这就是恒星诞生的故事。

无论是小质量恒星还是大质量恒星，一旦在其中心区域点燃氢核聚变，就会进入主序星的阶段。

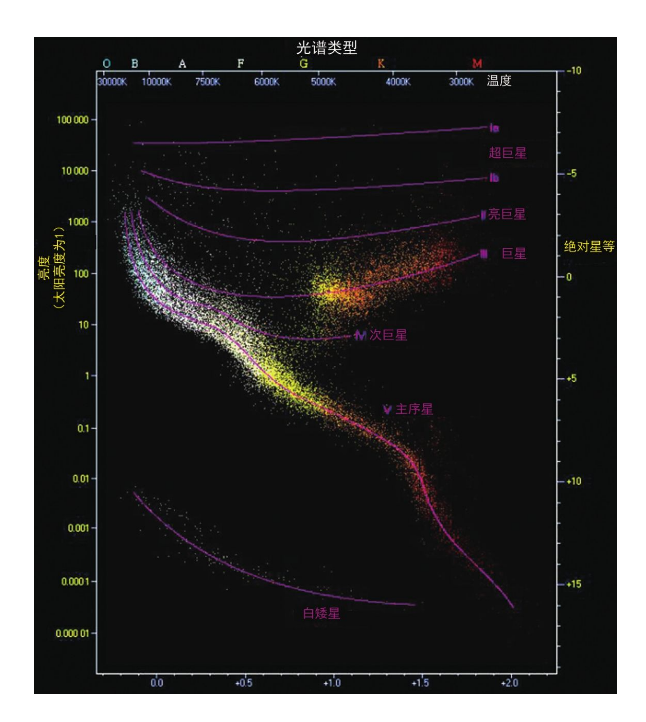
为了介绍什么是主序星，我得先给你科普一点天文学背景知识。

在20世纪初，丹麦天文学家埃纳尔·赫茨普龙（Ejnar Hertzsprung）和美国天文学家亨利·罗素各自独立地发明了一种研究恒星的重要工具，也就是所谓的赫罗图。

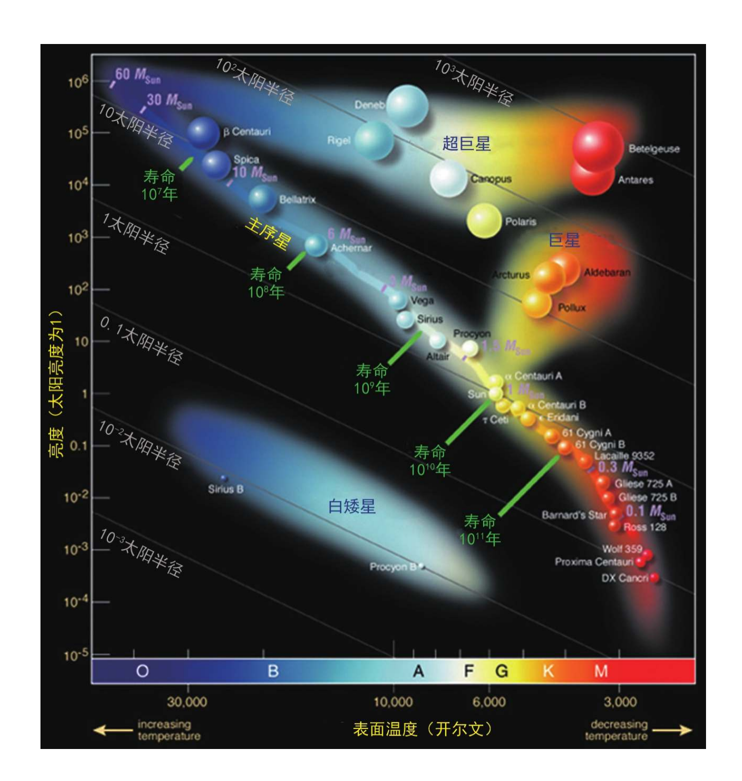




赫罗图是一个给恒星分类的二维直角坐标系，其横坐标代表恒星的表面温度，而纵坐标则代表恒星的绝对亮度（绝对亮度是假定把天体放在离地球32.6光年远的地方，所测得的亮度）。根据表面温度，恒星可以分为O、B、A、F、G、K、M七类。其中O型恒星的温度最高，超过30000开尔文，主要发出蓝白光；而M型恒星的温度最低，介于2400开尔文到3700开尔文，主要发出橙红光。而根据绝对亮度，按由亮到暗的顺序，恒星又可以分为超巨星、亮巨星、巨星和矮星。




后来人们发现，包括太阳在内的绝大多数的恒星，都分布在赫罗图中一条从左上角延伸到右下角的对角线上（即赫罗图主序对角线）。赫罗图主序对角线上的所有恒星，其表面温度都与其绝对亮度呈正相关。这些位于赫罗图主序对角线上的恒星，就是主序星（除了主序星以外，还有两个恒星聚集区域。一个位于赫罗图的右上角，称为红巨星；另一个位于赫罗图的左下角，称为白矮星）。




现在我们知道，天上绝大多数的恒星都是主序星。那么，主序星的本质是什么呢？最早揭开这个谜团的人，是英国大天文学家爱丁顿。



爱丁顿无疑是20世纪最伟大的天文学家之一。他一生中最有名的研究工作是基于1919年的日全食观测，验证了爱因斯坦的广义相对

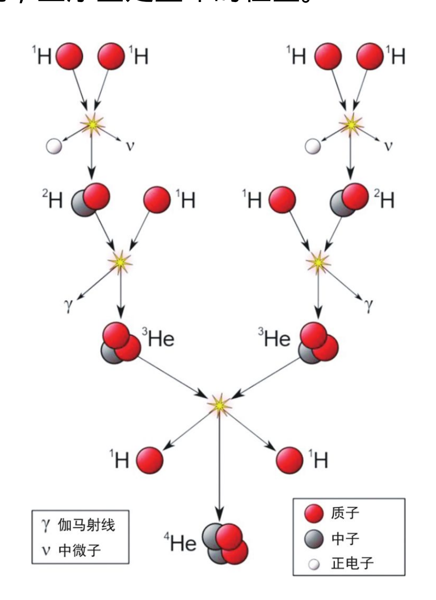
论。不过，这只是他在公众间知名度最高的工作，而不是他学术生涯的顶点。

真正奠定爱丁顿江湖地位的，是他在1920年发表的一篇名为《恒星内部结构》的论文。在这篇论文中，爱丁顿提出了一个最核心的问题：恒星到底靠什么来阻止自身的引力塌缩？正是这个问题，为人类揭开了恒星世界的神秘面纱。

爱丁顿对此问题的答案是，靠发生在恒星中心区域的氢核聚变。

氢核聚变会把4个氢原子核聚合成1个氦原子核，并释放大量的能量（此过程的能量转化率为7‰，比烧煤的能量转化率要高上百万倍）。这些能量可以产生方向向外的辐射压，进而与恒星受到的方向向内的引力达到平衡。正因为如此，恒星才可以持续稳定的存在。

依靠氢核聚变来对抗自身引力的恒星就是主序星，这就是主序星的本质。也就是说，主序星是盛年的恒星。




但是，一颗恒星中心区域的氢“燃料”并不是无穷无尽的。早晚有一天，恒星中心区域的氢燃料将会消耗殆尽，从而让氢核聚变中止（太阳中心的氢燃料还能再烧上50亿年。而质量是太阳10倍的恒星，只能再烧上几千万年）。到那时，恒星就会告别自己的盛年时期，迈

向暮年时代。

由于中心区域的氢燃料已经消耗殆尽，迈向暮年的恒星将在引力的作用下开始收缩。恒星的收缩会让它的温度整体升高。如此一来，原本温度较低的恒星外围的氢壳层，就可以突破核反应的临界温度，进而点燃氢核聚变。也就是说，氢核聚变会转移到恒星的外围区域。这样一来，恒星外围的氢壳层就不会再收缩，而是转为膨胀，从而让恒星的亮度大大超过之前的主序星阶段。而恒星外围氢壳层的膨胀，又会让它的温度下降，从而发出红光。

另一方面，恒星中心区域的氦壳层（氦是由之前中心区域的氢核聚变产生的）还在继续收缩，从而让核心温度不断升高。当核心温度超过1亿摄氏度的时候，就可以点燃氦核聚变，产生碳和氧元素，并释放大量的能量。

当中心区域的氦核聚变被点燃的时候，就能与引力达成新的平衡。换句话说，靠着中心区域的氦核聚变的支撑，迈入暮年的恒星将重新达到稳定的状态。此时，对于远处的观测者来说，这颗恒星将呈现出亮度大、温度低、发红光的特征。这就是所谓的红巨星。

也就是说，红巨星是暮年的恒星。

但红巨星中心区域的氦燃料，也会消耗殆尽。此后的恒星，就会迈向死亡。而小质量恒星和大质量恒星的命运，将出现分叉。

像太阳这样的小质量恒星，会有一场比较平淡的葬礼。它会抛出外围的氢壳层，形成被称为“行星状星云”的发光气体云；这些行星状星云最后会逐渐消散，成为星际介质的一部分。

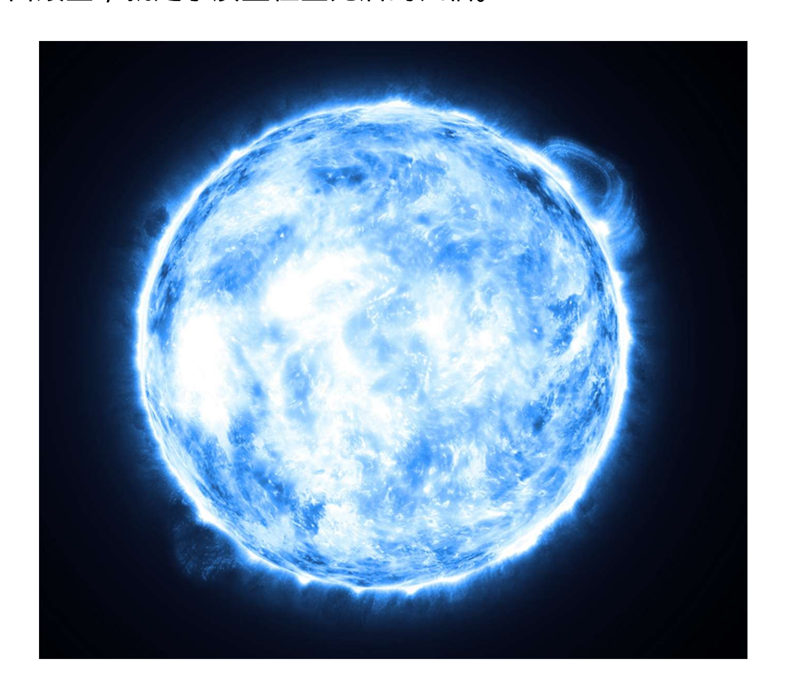
当所有的外围气体都被抛掉以后，由碳元素和氧元素构成的恒星内核就会暴露出来。这个内核还会继续塌缩。但由于质量不足，塌缩引起的温度升高始终无法点燃碳核聚变。最终，当这个恒星内核被引力压缩到和地球差不多大小的时候，它内部的电子简并压力[2]就可以与引力达到平衡。

当电子简并压力与引力达到平衡以后，这个恒星内核就可以稳定地存在下去了。此时，对于远处的观测者来说，这个残存的恒星内核

将呈现出亮度小、温度高、发白光的特征。这就是所谓的白矮星。

值得一提的是，白矮星有一个质量上限，也就是太阳质量的1.44倍，称其为钱德拉塞卡极限。一旦超过这个钱德拉塞卡极限，电子简并压力就无法再对抗引力。换言之，超过钱德拉塞卡极限的白矮星根本就无法存在。

白矮星，就是小质量恒星死后的归宿。



另一方面，质量能达到太阳质量10倍以上的大质量恒星，会有一场非常盛大的葬礼，也就是所谓的超新星爆发。

不同于死去的小质量恒星，大质量恒星的内核会因为自身的引力塌缩而达到极高的温度。这样一来，它就可以依次点燃碳、氧、硅的核聚变，直到在恒星最中心的位置产生一个铁核。这就形成了图16所示的“洋葱”结构。

与之前所有核聚变截然不同的是，铁核聚变不但不能释放能量，反而会吸收大量的能量。换句话说，铁核就不可能再聚变了。

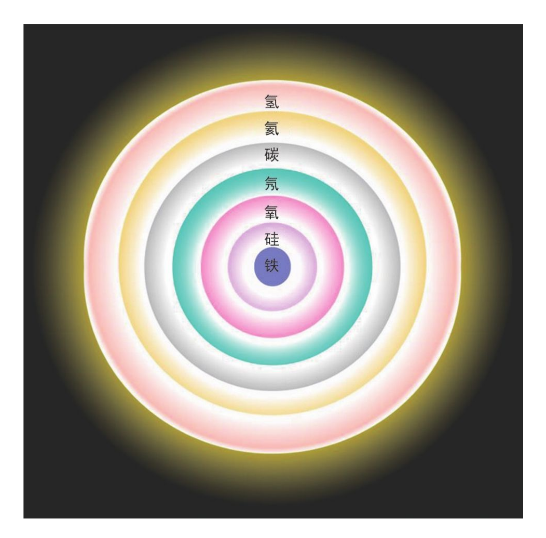
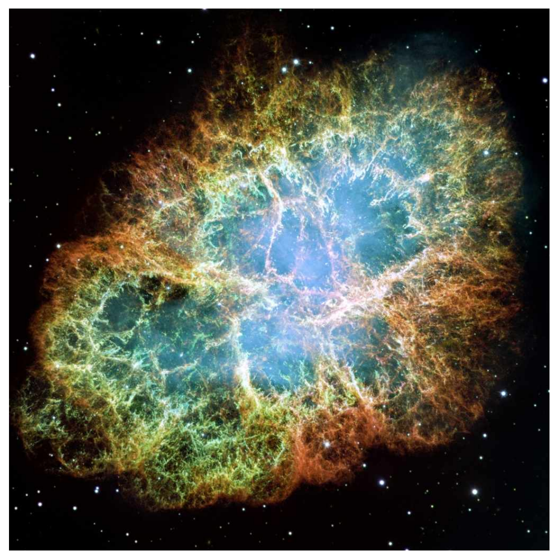



图16

在这种情况下，就连电子简并压力也无法再对抗恒星自身的引力。这意味着，引力会把铁核中的电子全部挤进原子核的内部。这些电子会与原子核内部的质子结合，变成中子。这就是所谓的恒星“中子化”过程。而恒星“中子化”的瞬间，会释放出海量的高能中微子（中微子是一种不带电荷、质量几乎为0的粒子，它也是宇宙中数量第二多的粒子）。这些高能中微子会向四面八方飞散，其实际效果就是一个中微子的大爆炸。这个中微子的大爆炸会把恒星的外层物质炸得四分五裂。由于发生了大爆炸的缘故，恒星的亮度能够达到平时的几千万倍，这就是超新星爆发。

超新星爆发是一场极端壮丽的宇宙烟花秀。在短短几十天内，这场烟花秀释放的能量，就能超过一颗恒星上百万年间释放的能量总和。正因为如此，一颗超新星的亮度就足以和一个星系相媲美。即使经历了上千年的岁月，超新星爆发的烟花秀依然能留下清晰可见的遗迹。其中最典型的例子，就是著名的蟹状星云。





超新星爆发后，会留下一个完全由中子构成的致密内核，这就是所谓的中子星。一般而言，中子星的半径约为10千米；而它的密度能达到水密度的400万亿倍。（一汤匙白矮星物质的质量，大概相当于一辆汽车；一汤匙中子星物质的质量，则大概相当于一座山。）不同于靠电子简并压力对抗自身引力的白矮星，中子星是靠中子简并压力来对抗自身引力的，这就是中子星的本质。

类似于白矮星，中子星也有一个质量上限，即太阳质量的3倍，称其为奥本海默极限。一旦超过这个奥本海默极限，中子简并压力也将无法对抗引力。这样一来，引力就会君临天下，最终把大质量恒星的内核压成一个黑洞。



黑洞是宇宙中最恐怖的监狱。它最核心的特征是：其逃逸速度能达到光速。正因为如此，只要进入了这个监狱的围墙（即黑洞的事件视界），就连宇宙中速度最快的光也不可能逃出它的魔掌[3]。

中子星和黑洞，就是大质量恒星死后的归宿。

我们已经讲完了恒星的一生。由于宇宙版本的马太效应，初始的宇宙密度涨落会逐渐演变成恒星。恒星在经历了盛年的主序星阶段和暮年的红巨星阶段后，会迈向死亡。小质量恒星的葬礼是行星状星云，然后留下一颗白矮星。大质量恒星的葬礼是超新星爆发，然后留下一颗中子星或一个黑洞。（还有一些质量最大的恒星，会直接塌缩成黑洞。）

但恐怖的是，最新的天文观测表明，在宇宙中能发光的所有恒星，其质量只占宇宙总质量的5%。这意味着，宇宙的真正主宰，其实是我们根本看不到的宇宙黑暗面，也就是所谓的暗物质和暗能量。

那么，人类是如何发现暗物质和暗能量的呢？

欲知详情，请听下回分解。



---

(1)宇宙版本的马太效应是英国天文学家詹姆斯·金斯（James Jeans）发现的，其学术名称是金斯不稳定性。

(2)当距离很近时，一个电子会对另一个电子产生排斥力，这就是电子简并压力。对它的物理图像感兴趣的读者，可以参阅我之前写的《宇宙奥德赛：穿越银河系》一书的2.3节。

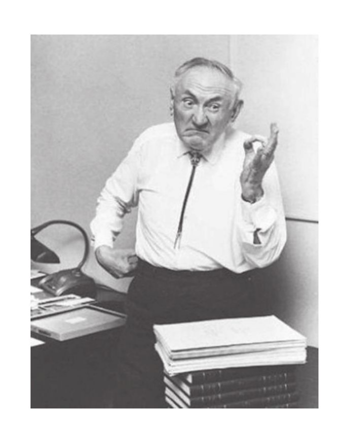
(3)对黑洞的物理图像感兴趣的读者，可以参阅我之前写的《宇宙奥德赛：穿越银河系》一书的7.2节。

# 11　暗物质与暗能量

上节课的结尾，我们提出了这样一个问题：人类是如何发现暗物质和暗能量的呢？

事实上，无论是暗物质还是暗能量的发现，都是足以载入物理学和天文学史册的重大历史突破。接下来，我就来讲讲人类发现暗物质和暗能量的故事。

暗物质的故事，得从一匹“独狼”讲起，他就是瑞士天文学家弗里茨·兹威基（Fritz Zwickg）。



兹威基是苏黎世联邦理工大学的博士，也就是爱因斯坦的校友。20世纪20年代，他移居美国，任教于加州理工学院，并在威尔逊山天文台做兼职研究员。很快地，他就成了人们眼中的怪胎。

兹威基生性粗鲁，喜欢骂自己看不上的人是“混球”。由于怕别人听不懂，还总要在后面补充一句：“混球就是具有球对称性、无论从哪个方向看都是混蛋的人。”那么，兹威基到底看不上哪些人呢？答案是加州理工学院和威尔逊山天文台的几乎所有人。

兹威基粗暴的性格让他树敌甚多。后来有一群人忍无可忍，给加州理工学院院长罗伯特·密立根（Robert　A.Millikan）写过一封联名信，强烈要求开除兹威基这个“恼人的小丑”。

但密立根没有同意。他在回信中写道："我知道兹威基是个疯子，但他在科学上提出了很多富有革命性的疯狂点子。万一这些点子里，有一两个是对的呢？"

实际上，兹威基总共对了4个，分别是超新星、中子星、引力透镜，以及接下来要重点介绍的暗物质。

20世纪30年代，兹威基开始研究后发座星系团。为了测量这个星系团的质量，他采用了两种截然不同的方法：光度学方法和动力学方法。光度学方法通过测量星系团发出的光的亮度，来估算星系团中发光物质（即恒星）的质量；动力学方法则通过测量星系团边缘的天体的运动速度，来计算整个星系团的总质量。



兹威基最后发现，用动力学方法测出的星系团总质量，是用光度学方法测出的发光物质质量的400倍。换言之，在星系团中存在的绝大多数的物质，我们都是看不见的。（以今天的眼光来看，兹威基的这个测量结果是错误的。他之所以会搞错，是因为他在估算过程中使用了当时流行的、但实际上是错误的哈勃常数。）

为了解释这个诡异的观测结果，兹威基提出了一个相当"疯狂"的点子：星系团中存在着一种看不见的物质，也就是所谓的暗物质（严

格地说，兹威基并不是提出“暗物质”这个名词的第一人。但他最早用天文观测证明，暗物质的存在是一个非常现实的问题）。

需要强调的是，这里的“暗”并不是指黑暗，而是指透明。黑暗的物质会彻底吸收光，而透明的物质则会直接无视光。换句话说，暗物质根本不会与光发生任何相互作用。这样一来，光就可以毫无障碍地直接穿过暗物质，而不会被暗物质反射。因此，我们永远无法直接看到暗物质。

但是在科学领域，远远超越时代的先知往往会变成“先烈”。兹威基就没能逃脱这样的宿命。在长达40年的时间里，兹威基提出的这个暗物质理论，一直无人问津。

直到20世纪70年代，另一个人的横空出世才让暗物质得到了普遍的承认。此人就是美国天文学家薇拉·鲁宾（Vera Rubin）。

类似于我们之前介绍过的现代宇宙学之母勒维特，鲁宾在追求科学的道路上，也遭遇了很多歧视和不公。举个例子,在高中毕业那年，她向一所大学提出了申请，想去那里学习天文学专业。但是，招生面试官觉得女性不适合研究科学，竟试图引导她去学习更为“淑女”的美术专业。后来，这成了鲁宾朋友圈中的一个笑话。只要她在工作中遇到了挫折，就一定会有人问：“你是否考虑过画画的职业？”本科毕业后，成绩优异的鲁宾满怀憧憬地申请了普林斯顿大学天文系的研究生项目。结果她被告知，普林斯顿大学天文系根本不招女生（直到1975年，普林斯顿大学天文系才开始招收女生）。



但种种歧视和不公不但没打倒鲁宾，反而塑造了她强悍的个性。

1954年，鲁宾在乔治敦大学拿到了博士学位，随后成为了卡内基科学研究所的首位女研究员。1965年，她得到许可，可以用帕洛玛山天文台的大型望远镜进行天文观测。这也让她成了历史上第一个获此殊荣的女天文学家。

但到了帕洛玛山天文台以后，鲁宾发现了一个问题：这里根本就没有女卫生间。于是她就把一张纸剪成了短裙的形状，并贴在了一个男卫生间的门上。然后她就守在那个卫生间的门口，把所有想去那里上厕所的男人都赶跑。从那以后，帕洛玛山天文台就有了女卫生间。

20世纪60年代末，鲁宾开始与她的同事肯特·福特（Kent Ford）合作，研究一个当时很不起眼的领域：测量星系的旋转速度。



鲁宾选定的研究对象，是我们之前游览过的仙女座星系。在观测开始前，鲁宾和福特一致认为，他们肯定会看到这样的景象：离星系中心越近的恒星，绕星系中心公转的速度就越大；离星系中心越远的恒星，绕星系中心公转的速度就越小。这也是我们在太阳系中看到的景象：离太阳越近的行星，其公转速度就越大；离太阳越远的行星，其公转速度就越小。

但最后的观测结果让鲁宾和福特都大吃一惊。他们惊愕地发现：恒星的实际公转速度竟然是一个常数，与恒星到星系中心的距离无关。

图17就展示了鲁宾和福特的发现。此图中的横轴表示恒星到星系中心的距离，纵轴则表示恒星的运动速度。鲁宾和福特本以为他们会

看到恒星的公转速度随距离的增大而降低，也就是图中的虚线。但实际上，他们发现恒星的公转速度是一个常数，与距离无关，也就是图中的实线。这条后期逐渐变平的实线，就是著名的星系旋转曲线。

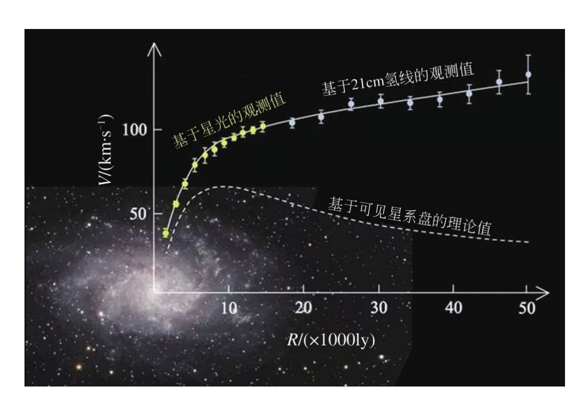



图17

起初，鲁宾和福特还以为这种诡异的结果是仙女座星系独有的。但后来，他们又研究了200多个星系，并发现了所有的星系都有一条相同的、后期逐渐变平的旋转曲线。这意味着，恒星公转的速度与它到星系中心的距离无关，这是一条适用于所有星系的普遍规律。

为什么说这个结果非常诡异呢？原因在于，如果星系边缘恒星的公转速度不随距离的增大而降低，那么它就可以挣脱星系引力的束缚，飞到遥远的太空中去。而随着星系边缘的恒星不断被剥离，整个星系也将土崩瓦解。但实际情况是，星系可以非常稳定地存在几十亿年。这到底是怎么回事呢？

唯一合理的解释，就是星系总质量远远大于我们看到的发光物质的质量。换句话说，在星系中必须存在大量的看不见的物质，它们提供的额外的引力牢牢地束缚住了星系边缘的恒星。也只有这样，整个星系才不会土崩瓦解。

星系中大量存在的这种看不见的物质，就是兹威基40年前预言的暗物质。根据鲁宾的估算，对于所有星系而言，其中包含的暗物质的质量，至少是能发光的恒星总质量的5～6倍。

1975年，鲁宾在美国天文学会的年会上报告了自己的发现。她指

出：所有星系的旋转曲线都有后期变平的现象，这说明了所有的星系中都存在着大量的暗物质。这是人类历史上首次发现暗物质存在的确凿证据。

而暗物质的存在，后来也得到了其他天文观测（如引力透镜和星系团并合）的证实。

我们已经介绍了人类是如何发现暗物质的。接下来，就该讲暗能量了。

暗能量的故事能一直追溯到爱因斯坦。之前我们讲过，为了维持一个静态的宇宙，爱因斯坦在他的引力场方程中引入了一个宇宙常数项；这个宇宙常数项能产生斥力，从而与整个宇宙的引力达成平衡，并让整个宇宙保持静止。1931年，哈勃发现了宇宙在膨胀。这让爱因斯坦追悔莫及，宣称引入宇宙常数是他一生中"最大的错误"。

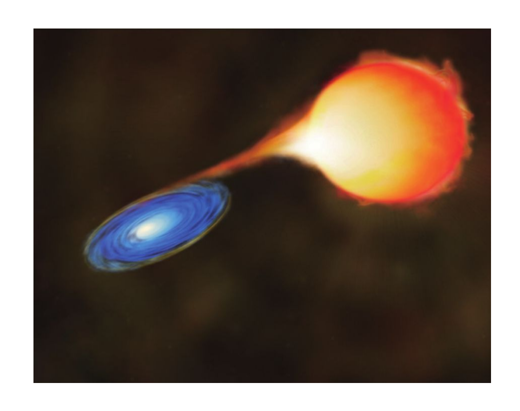
后来，有些人［如苏联大天文学家泽尔多维奇（Zel'dovich）］也曾试着拯救这个宇宙常数理论，但全都铩羽而归。直到20世纪90年代末，两个美国的天文观测组做出了一个划时代的重大发现，这才让爱因斯坦的宇宙常数王者归来。

那两个观测组的科学目标，是利用Ia型超新星测量宇宙的膨胀速率。

先介绍一下什么是Ia型超新星。宇宙中大多数的恒星都处于双星系统。在两颗互相绕转的恒星中，肯定有一颗会先死，并且变成一颗白矮星。随后，没死的那颗恒星也会迈向暮年时代，并变成一颗红巨星。这样一来，白矮星就可以从体积膨胀的红巨星那里吸积物质，形成一个宛如海底漩涡的吸积盘。一旦白矮星和吸积盘的总质量超过了钱德拉赛卡极限（太阳质量的1.44倍），就会引发一场巨大的核爆炸，从而让自身的亮度急剧增大。这场由白矮星吸积伴星物质所引发的大爆炸，就是Ia型超新星爆发。

因为所有的Ia型超新星爆发时所释放的总能量一定是太阳质量的1.44倍，所以可以近似地认为，Ia型超新星的绝对亮度固定不变。这样一来，就可以把Ia型超新星视为标准烛光，来进行距离测量。另一方面，Ia型超新星相对于地球的视向速度，可以用多普勒效应来测。

通过比较一批Ia型超新星的视向速度和它们到地球的距离，就可以确定宇宙的膨胀速度了。



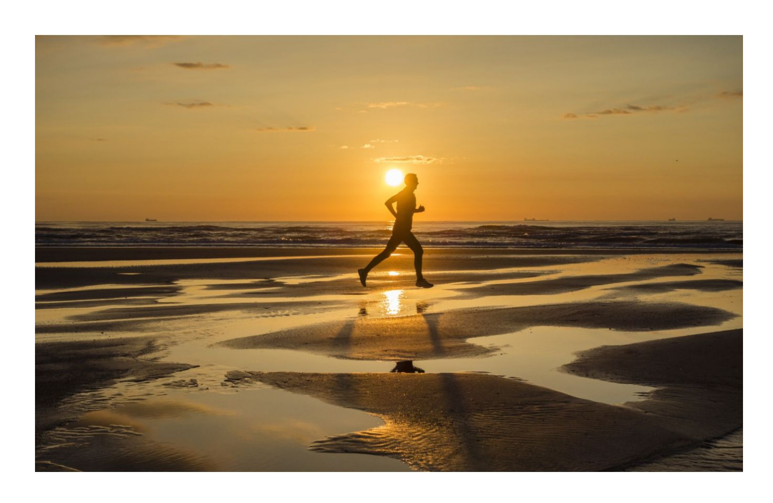
但这两个天文观测组的测量结果，让所有人都惊掉了下巴。他们的结果表明：宇宙不但在膨胀，而且在加速膨胀。

这到底是怎么回事呢？我们来打一个比方。

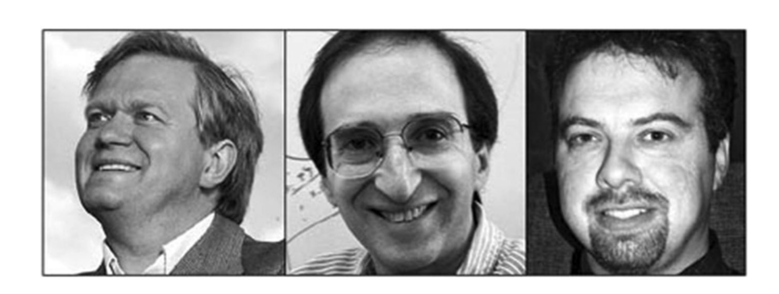
想象有一个田径运动员，他跑步的速度是恒定的10米/秒。现在，让他在逆风的环境下跑上10秒。10秒之后，我们再测量他所跑的距离。

按理说，由于逆风，这个运动员跑过的距离肯定不到100米。但实际的测量结果表明，他跑过的距离竟然远远超过了100米。这是怎么回事呢？唯一的可能是，运动员所处的环境根本不是逆风，而是顺风。

现在，让我们把这个奔跑的运动员想象成一个正在膨胀的宇宙。逆风意味着，引力的存在会让宇宙的膨胀减速；但实际的测量结果表明，宇宙的膨胀不但没有减速，反而在不断加速，这就是所谓的宇宙加速膨胀。



1998年，这两个天文观测组各发表了一篇论文，宣布他们发现宇宙正在加速膨胀。这是继哈勃发现宇宙膨胀以来最重大也最震撼的宇宙学发现，被《科学》杂志评为了当年的十大科学突破之首。这个发现也让三位美国科学家[索尔·珀尔穆特（Saul Perlmutter）、布莱恩·施密特（Brian Schmidt）、亚当·里斯（Adam Riess）]获得了2011年的诺贝尔物理学奖。



宇宙加速膨胀意味着：主宰整个宇宙的并不是引力，而是斥力（斥力就对应于运动员所处的顺风环境）。那么，这种神秘的斥力到底从何而来？

目前学术界最主流的观点是：这种斥力源于一种非常神秘的事物，也就是所谓的暗能量。

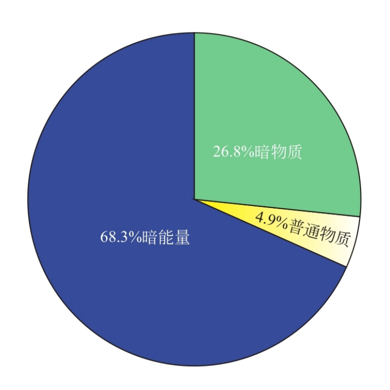
暗能量有三个最核心的特征。第一，它是透明的。也就是说，它不会与光发生任何相互作用，因而永远也不会被看到，所以才叫“暗”。第二，它会产生斥力。因此它与物质存在着本质上的不同，

所以才叫“能量”。第三，它在宇宙中均匀分布，不会聚集成团。事实上，它是一种源于真空的能量，藏在我们每个人的体内和身边。

那我们为何在日常生活中完全感受不到暗能量的存在呢？因为它的密度太小了，每立方厘米内的质量还不到$10^{-29}$克。如果把100多个地球内包含的暗能量都加在一起，也只有区区1克。因此，我们在宏观尺度上完全无法感知暗能量的存在。但是放眼整个宇宙，暗能量聚沙成塔，变成了主宰整个宇宙的力量（最新的天文观测表明，暗能量目前占宇宙总物质组分的68.3%）。

暗能量到底是什么呢？20多年后的今天，人类对此依然知之甚少。

到目前为止，物理学家们已经提出了成百上千种暗能量模型。但目前最受天文观测青睐的，依然是爱因斯坦在100多年前提出的宇宙常数模型。这个宇宙常数模型说的是：源于真空的暗能量的能量密度，永远都是一个常数，不会随时间推移而发生改变。




基于暗物质和暗能量的发现，科学家们构造了一个“标准宇宙模型”，也叫“ΛCDM模型”。它说的是，我们的宇宙正由暗能量（即宇宙常数Λ）和冷暗物质（即运动速度缓慢的暗物质，cold dark matter，简称CDM）统治。其中暗能量占宇宙总物质组分的68.3%，冷暗物质占26.8%。

或许你会好奇，暗能量到底有什么用。答案是，它将主宰宇宙的最终命运。那么，暗能量将会如何主宰宇宙的最终命运？

欲知详情，请听下回分解。

# 12 宇宙的终极命运

上节课的结尾，我们提出了这样一个问题：暗能量将会如何主宰宇宙的最终命运？

在回答这个问题前，我们得先讲讲暗能量为什么能主宰宇宙的命运。

我们之前讲过，暗能量目前占宇宙总物质组分的68.3%。随着宇宙的不断膨胀，宇宙中物质的密度将不断减小（随着宇宙的膨胀，宇宙的体积将不断增大，但其中包含的物质总量不变，所以密度就会不断减小）。但根据爱因斯坦的宇宙常数理论，暗能量的密度只取决于真空的性质，而与宇宙的膨胀无关。因此，暗能量的密度始终是一个常数。这意味着，随着时间的推移，暗能量在宇宙总物质组分中所占的比例将不断提高，并最终趋向于100%。所以，宇宙的命运必将由最终占比100%的暗能量主宰。

接下来，我要介绍宇宙有哪些可能的命运。在暗能量发现之前，人类普遍认为，宇宙有三种可能的命运，即“大挤压”“大反弹”和“大冻结”。为了讲清楚“大挤压”“大反弹”和“大冻结”的含义，我要先做一个类比。

想象有一个乒乓球，被你用力地抛向空中。那么这个乒乓球就有三种可能的结局。

第一种结局，乒乓球飞行的速度不够快，被地球引力拉了回来,随后一头栽在地上，再也弹不起来。

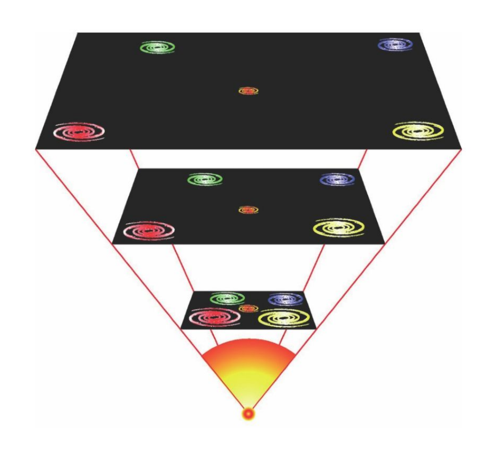
第二种结局，乒乓球飞行的速度不够快，被地球引力拉了回来,接着被地面反弹，然后又被地球引力拉回。如此弹起、落下、弹起、落下，不断循环。

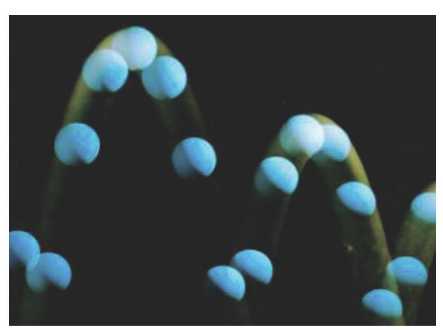
第三种结局，乒乓球飞行的速度足够快，从而彻底挣脱地球引力的束缚，飞向太空，一去不复返。

现在，把乒乓球的飞行想象成宇宙的膨胀，把地球的引力想象成整个宇宙的引力，这样就可以把上述的三种结局与宇宙的三种命运一一对应。

第一种结局对应“大挤压”。它说的是，宇宙将来会由膨胀转为收缩，并最终将其中包含的所有物质都挤压进一个体积无穷小、密度无穷大的时空奇点。

第二种结局对应“大反弹”。它说的是，宇宙将来会由膨胀转为收缩，而收缩到足够小的时候又会被反弹，从而重新开始膨胀。这样一来，宇宙就会在不断的膨胀与收缩中，循环往复。





但是暗能量的发现，几乎宣判了“大挤压”和“大反弹”理论的“死刑”。因为，乒乓球已经不会再落回地面了；它将在暗能量所产生斥力的推动下，加速飞离地球。因此，宇宙将走向第三种结局，即挣脱

引力束缚，永远膨胀下去（也有极少数的暗能量理论认为，宇宙还是有可能在遥远的未来由加速膨胀转为最终收缩）。换言之，宇宙的最终命运将是“大冻结”。

“大冻结”意味着，宇宙将迎来一个黑暗、寒冷、孤独的死亡。在此过程中，有以下几个标志性事件。

（1）在目前的宇宙中，既有恒星在死亡，也有恒星在诞生。但早晚有一天，宇宙中所有的恒星都会死亡，并且不会再有新的恒星诞生。此后，宇宙就将陷入永恒的黑暗。

（2）宇宙加速膨胀会让室女座超星系团（也就是我们住的这个“省”）以外的所有星系，都离我们越来越远，直到再也无法看见。换句话说，宇宙加速膨胀会不断扩大宇宙之海的尺度；被引力束缚、面积无法扩大的超星系团，将变成漂浮在这片海里的宇宙孤岛。

（3）随着动能的耗尽，所有的人造地球卫星最后都会落回地球。类似地，随着动能的耗尽，所有的天体最后都会落入超星系团中心的超大质量黑洞。到那时，所有的宇宙孤岛都会变成无比巨大的黑洞，像一个个盘踞在宇宙中的可怕怪物。

（4）黑洞依然不是终点。随着宇宙的膨胀，宇宙微波背景的温度将不断降低，最终会低于所有黑洞的温度。此后，黑洞就会开始蒸发（这就是所谓的霍金辐射），变得越来越小。早晚有一天（一般认为，至少要花$10^{1000}$年），宇宙中所有的黑洞都会蒸发殆尽。到那时，宇宙中的万事万物都会烟消云散。

黑暗、寒冷、几乎空无一物，这就是宇宙“大冻结”的最终结局。

但在20世纪末，有人发现“大冻结”并非宇宙唯一可能的命运。此人就是美国物理学家罗伯特·考德威尔（Robert Caldwell）。

1999年，考德威尔提出了一个全新的暗能量模型。当时恰好有一部好莱坞大片在热映，那就是《星球大战1：幽灵的威胁》。为了向这部大片致敬，考德威尔用幽灵的英文单词phantom来给自己的模型命名，其中文译名是幻影暗能量。



但是考德威尔写的这篇提出幻影暗能量的论文，却遭到了学术界的围剿。它遭到了数名审稿人的刁难，直到三年后才得以正式发表。为什么大家都不喜欢这篇论文呢？原因在于，它揭示了一种匪夷所思的可能性：暗能量的密度会随着时间的推移而不断变大。

潘多拉魔盒就这样打开了。一场灾难也随之降临。

我来解释一下，这到底意味着什么。我们所熟悉的世界，其稳定是靠引力维系的。而且，对于一个引力束缚系统（如行星、恒星、星系和星系团）而言，引力的大小是固定的，不会随时间的推移而发生改变。

但是充斥在宇宙的每个角落、并且能产生斥力的暗能量就不同了。特别是这个幻影暗能量，其能量密度会随着时间的推移而不断变大。这就意味着，它产生的斥力也会越来越大。

目前，暗能量的密度还不到$10^{-29}$克/立方厘米，所以我们完全感受不到它发出的斥力。但要是暗能量产生的斥力能随着时间的推移而不断变大，早晚有一天，它将超过所有引力，从而破坏原本由引力维系的整个世界的稳定。换句话说，到时宇宙中所有的结构，无论是银河系、太阳系、地球还是我们自身，都会被幻影暗能量从内部撕碎。幻影暗能量从内部撕碎一切的这个恐怖末日景象，就是所谓的宇宙“大撕裂”。



宇宙“大撕裂”到底是一个怎样的景象？2012年，我与4位同事合作，写了一篇研究宇宙最终命运的论文，并得到了几十家中外媒体的报道。我们的研究结果表明，宇宙“大撕裂”确实有可能发生。在最坏的情况下，宇宙甚至有可能在167亿年后就遭遇毁灭。

接下来，我就基于这篇论文，为你播放一部关于宇宙“大撕裂”的末日影片。

假设宇宙大撕裂发生在公元167亿年12月31日的24：00。公元167亿年与现在最大的不同是，天上的星星早已全部消失。除此以外，在最后一年大多数的时间里，我们并不会感受到任何的异常。

但到了10月31日，冥王星会突然消失。随后，海王星、天王星、土星、木星和火星，也会一个接一个地神秘失踪。

到了12月26日，月球也离家出走了；它挣脱了地球引力的束缚，像脱缰的野马一样，消失在了太空的深处。

真正恐怖的事情发生在12月31日的午夜。那天晚上的23：32，太阳死后留下的那颗白矮星，会突然分崩离析。到了23：44，地球也突然土崩瓦解。在末日到来前的$10^{-17}$秒，就连原子都会被幻影暗能量的强大斥力撕碎。然后就是“大撕裂”的时刻。这时幻影暗能量将君临天下，彻底摧毁宇宙中的一切。整个宇宙，甚至包括时间本身，都会在这一刻走向终结。

美国桂冠诗人罗伯特·弗罗斯特（Robert Frost）在他的名作《火与冰》中写下了这样的诗句：“有人说世界将终结于火，有人说是冰。”这恰好对应宇宙可能面临的两种最终命运：“大撕裂”和“大冻

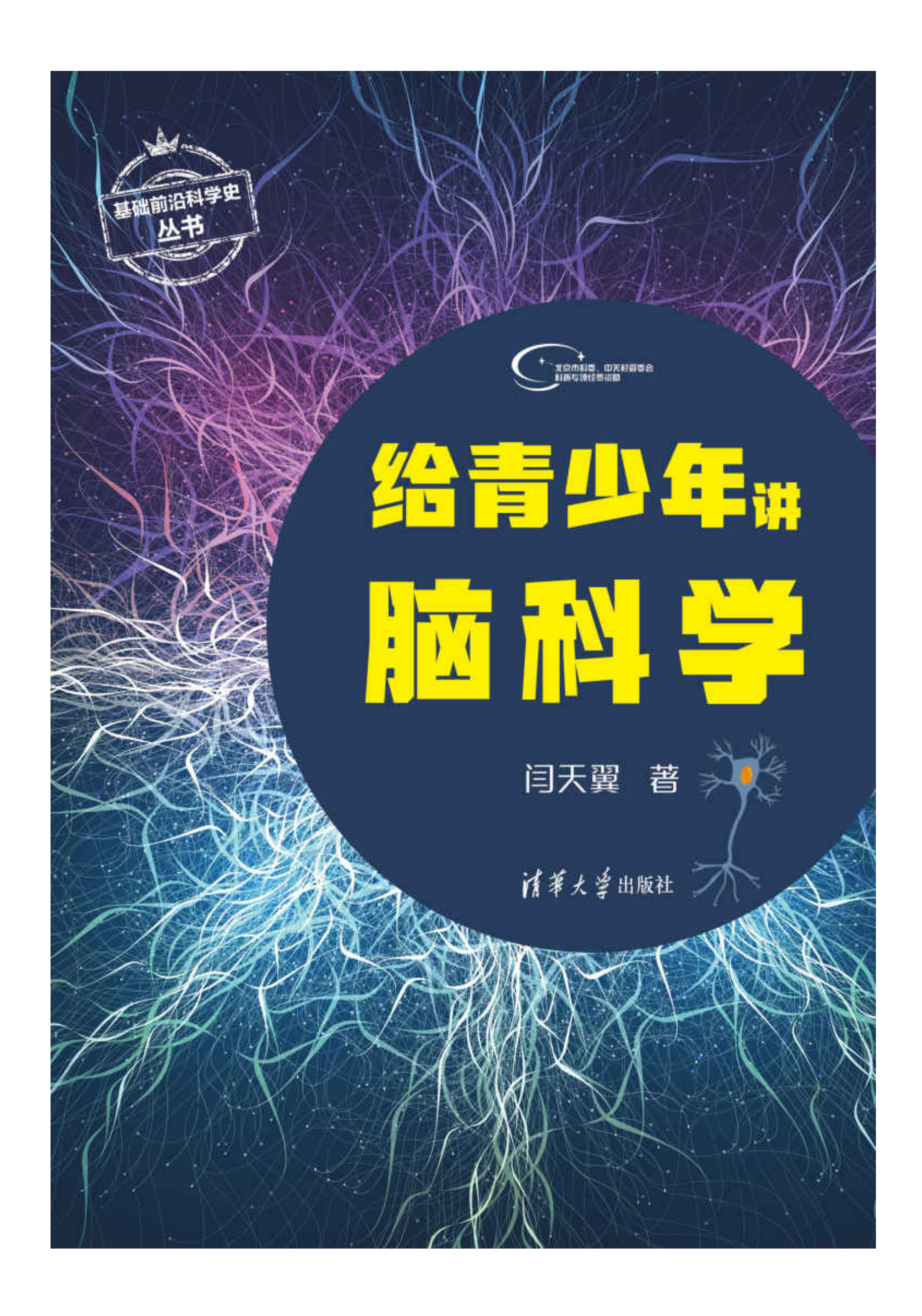
结”。所以宇宙是一首最典型的“冰与火之歌”。

但不管是冰的结局还是火的结局，宇宙最后都会变成一个黑暗、寒冷、空无一物的地方。正所谓“好一似食尽鸟投林，落了片白茫茫大地真干净”。



# 图片来源




基础前沿科学史
丛书
给青少年讲
脑科学
闫天翼 著
清華大學出版社

基础前沿科学史丛书

# 给青少年讲脑科学

闫天翼　著

清華大学出版社
北　京

图书在版编目（**CIP**）数据

给青少年讲脑科学／闫天翼著．—北京：清华大学出版社，2022.10

（基础前沿科学史丛书）

ISBN 978-7-302-62004-4

I．①给…　II．①闫…　III．①脑科学－青少年读物　IV．①Q983-49

中国版本图书馆CIP数据核字（2022）第187031号

责任编辑：刘　杨

封面设计：意匠文化·丁奔亮

责任校对：王淑云

责任印制：宋　林

出版发行：清华大学出版社

网　　址：http://www.tup.com.cn, http://www.wqbook.com

地　　址：北京清华大学学研大厦A座

邮　　编：100084

社总机：010-83470000

邮　　购：010-62786544

投稿与读者服务：010-62776969，c-service@tup.tsinghua.edu.cn

质量反馈：010-62772015，zhiliang@tup.tsinghua.edu.cn

印装者：三河市龙大印装有限公司
经　　销：全国新华书店
开　　本：165mm×235mm
印　　张：10.25
字　　数：111千字
版　　次：2022年12月第1版
印　　次：2022年12月第1次印刷
定　　价：55.00元

产品编号：097619-01

# 丛书序　给面向青少年的科普出版点一把新火

2022年是《中华人民共和国科普法》通过的第20年，在这样一个对科普工作意义不凡的年份，由北京市科学技术委员会（以下简称市科委）发起，清华大学出版社组织的“基础前沿科学史丛书”正式出版了。这套书给面向青少年的科普出版点了一把新火。

2022年9月4日，中共中央办公厅、国务院办公厅印发《关于新时代进一步加强科学技术普及工作的意见》，进一步强调“科学技术普及是国家和社会普及科学技术知识、弘扬科学精神、传播科学思想、倡导科学方法的活动，是实现创新发展的重要基础性工作”。科学技术普及是科技知识、科学精神、科学思想、科学方法的薪火相传——是“薪火”，也是“新火”。

市科委搭台，出版社唱戏，这套书给面向青少年的科普图书出版模式点了一把新火。市科委于2021年11月发布了“创作出版‘基础前沿科学史’系列精品科普图书”的招标公告，明确要求中标方在一年的时间内，以物质科学、生命科学、宇宙科学、脑科学、量子科学为主题，组织“基础前沿科学史”系列精品科普图书（共5册）出版工作；同步设计制作科普电子书；通过网络媒体对图书进行宣传推广等服务内容。这些服务内容以融合出版为基础，以社会效益为初心。服务内容的短短几句话，每一句背后都是特别繁复的工作内容。想在一年的时间内，尤其是在2022年新冠肺炎疫情期间，完成这些工作的难度可想而知，然而秉承“自强不息，厚德载物”的清华大学出版社的出版团队做到了。

中国科学家，讲好中国故事，这套书给面向青少年的科普图书选题内容点了一把新火。中国特色社会主义进入新时代，新一轮科技革命和产业变革正在深入发展，基础前沿科学改变着人们的生产生活方式及思维模式。《中华人民共和国国民经济和社会发展第十四个五年规划和2035年远景目标纲要》提出：在事关国家安全和发展全局的基

础核心领域，制定实施战略性科学计划和科学工程。物质科学、生命科学、宇宙科学、脑科学、量子科学等领域，迫切需要更多人才参与研究，而前沿科学人才的建设培养，要从青少年抓起。这5本书的作者都是中国本土从事相关专业领域工作的科学家，这5本书都是他们依托自己工作进行的原创性工作。虽然内容必然涉及科学史的内容，但中国科学家尤其是近些年的贡献也得到了充分展示。

初心教育，润物无声，这套书给面向青少年的科普图书科普创作点了一把新火。习近平总书记提出：科技创新、科学普及是实现创新发展的两翼，要把科学普及放在与科技创新同等重要的位置。因此，针对前沿科技领域知识的科普成为重点。如何创作广受青少年欢迎的优秀科普图书，充分发挥科普图书的媒介作用，帮助青少年树立投身前沿科学领域的梦想，是当前科普出版工作的重点之一，这对具体的科普创作方法提出了要求。这套书，看得出来在创作之初即统一了整体创作思路，在作者进行具体创作时又保持了自己的语言习惯和科普风格。这套书充分体现了，面向青少年的科普图书创作，应该循序渐进，张弛有度，绘声绘色，娓娓道来，以科学家的故事吸引他们，温故科学家的研究之路，知新科学家的科研理念，以科学精神润物细无声。

靡不有初，鲜克有终。2022年10月16日，习近平总书记在中国共产党第二十次全国代表大会报告中强调“教育、科技、人才是全面建设社会主义现代化国家的基础性、战略性支撑”。且将新火试新茶，诗酒趁年华。希望清华大学出版社的这套“基础前沿科学史丛书”为广大青少年推开科学技术事业的一扇门，帮助他们系好投身科学技术事业的第一粒扣子，在全面建设社会主义现代化强国的新征程上行稳致远。

中国工程院院士<br>清华大学教授



# 前言　欢迎来到大脑的世界！

“大脑是你最重要的器官”——这是由大脑告诉你的。

我们为什么会认识苹果？为什么会知道口渴？是什么让我们保持思考和学习？又为什么我们会有喜、怒、哀、惧、爱、恶这些情绪与情感？各位同学有没有思考过这些问题呢？

人类从很久之前就开始关注“大脑”了。无论是西方还是中国，起初人们都将心脏视为记忆和思考的器官，这种认知一直到古希腊时代才逐渐改变。公元前4世纪，古希腊时代的先哲认为“思维、情感、智慧皆来自于大脑，大脑参与对环境的感知”。1795年，人们才确定人类的思维来自大脑。到了18世纪末，随着对人体的进一步探索和科技的发展，人类对大脑的大体解剖已经有了较为细致的描述，这奠定了“不同脑功能定位于不同的脑回”的理论基础，为脑功能定位研究开创了新时代。当时间来到20世纪，科学家发现，尽管特定的大脑区域负责某项独立的功能，但这些区域组成的网络以及它们之间的相互作用才是人类表现出整体、综合行为的原因，即大脑是一个活跃的、动态的系统。进入21世纪以来，脑科学研究呈现大发展的态势。科学家们以“脑探知、脑保护和脑创造”为目的，通过脑成像学、分子生物学、解剖生理学等研究方法，从已知的宏观层面进入介观层面再到微观层面，认识和了解大脑的结构和功能，进而开发和模拟大脑，实现创造和融合大脑。

脑是人体最复杂的器官，是人体一切行为、思维、决策和感觉的司令部。然而，目前人类对大脑的了解尚处于初级阶段。更好地了解大脑的构造和功能，对教育、医疗乃至人类发展意义重大。正是基于那些来自医疗、科研和技术的需求，人类脑计划应运而生。自2013年起，美国、欧洲、日本相继启动各自大型脑科学计划，全球参与脑计划的国家数量不断增加，它不仅仅是科技发展的信号，更代表了全球化科研资源的整合。中国于2021年正式启动了“脑科学与类脑科学研究”，即“中国脑计划”，提出了“一体两翼”战略，即以研究脑认知的神经原理为“主体”，以研发脑重大疾病诊治新手段和脑机智能新技术

为“两翼”。“十三五”规划提出强化脑与认知等基础前沿科学研究，将脑科学与类脑研究纳入“科技创新2030——重大项目”；“十四五”规划则明确提出瞄准脑科学等前沿领域。脑科学作为我国一个相对较新的研究方向，目前正处于积极发展阶段，此前全球已实施脑计划的经验和教训对我国进行脑计划也有所帮助。时至今日，理解脑的工作机制，对于重大脑疾病的早期预防、诊断和治疗，人脑功能的开发和模拟，创造以数值计算为基础的类脑智能，以及抢占国际竞争的技术制高点具有重要意义。相信在全球各具特色的脑计划共同协作下，人类对脑和疾病的认知将不断深入，并从中寻找到更为广泛的应用价值。

其实早在20年前读大学本科期间，我就对心理学非常感兴趣，但那时接触到的书籍更多的是在介绍脑的心理或生理基础，理论性很强，直白又枯燥。本科毕业之后，我选择继续攻读硕士和博士学位，在此期间，我接触了脑电、核磁等大脑信息解码技术，了解到大脑的编码、信息处理，甚至记忆与情感都可能被量化，想象与梦境都可能被再现……这使我笃定信念，要从事这个被科学界视为“皇冠上的明珠”的科研领域。经过了近20年的学习与研究、教学与实践，我越来越深入地认识到了脑科学的奇妙之处，也深知脑科学的探索需要几代人甚至几十代人共同努力，而我们也终将在不断进行的脑科学探索中推动人类科学与文明的进步。

面向脑科学的国际研究前沿，国家在脑基础科学、脑机智能等领域的重大需求，在生物医学工程、机械工程和计算机科学及其交叉领域，我所在的北京理工大学研究团队正在进行相关创新性基础和应用研究，培养领域内高精尖人才。我本人的研究领域所在学科方向为脑科学与神经工程，以脑基础科学、脑机智能技术研究为主线，涉及脑机制、脑模拟、脑康复领域的理论研究和仪器设备研发等工作。

在国家大力支持脑科学与类脑研究的背景下，越来越多的研究人员和团队加入到脑科学研究的浪潮中。本书整理和借鉴了前人的部分研究成果，结合笔者多年的研究经验，主要介绍了大脑研究的发展历程，大脑的基础知识，以及脑科学在实际生活中的应用。为了兼顾趣味性和科学性，本书使用了类比手法，将大脑的知识变成了大家在实际生活中的所见所闻，既可以激发青少年的阅读兴趣，同时还能掌握

相关的科学知识。

青少年是祖国的未来，希望这本书可以让广大青少年更加客观地了解大脑、认识大脑、理解大脑，激发青少年对脑科学研究的热情，新一代的脑科学研究力量可能就来自于各位热爱脑科学的同学们。最后，由于本人能力有限，难免存在缺点和不足之处，祈望读者批评指正。祝阅读和学习愉快！

闫天翼<br>2022.11

# 目 录

# 1　脑科学的前世今生

人脑被认为是自然界中最复杂、最高级、最精密的智能系统，揭示脑的奥秘已成为当代自然科学面临的巨大挑战之一。然而，对人脑认识和研究的历史却远比你想象得久远。现有证据表明，我们的史前祖先也许早就已经意识到了大脑在生命活动中的重要作用。

1865年，一位考古学家（名字已无从考证）在经过印加古城的时候，从一位女收藏家那里得到了一颗特别的头骨，如图1-1所示。这颗头骨的头盖骨部分有一个洞。这位考古学家认为，这个洞并不是常见的头部创伤，而是一个手术的结果——这个头骨的主人在手术后还短暂地存活了一段时间。但根据当时的医院对患者进行这样的脑手术存活率都较低的情况，大多数人认为医疗手段与技术更加落后的古印加人不可能完成这么复杂的手术。就这样过了7年，当人们在一个新石器时代遗址处发现了多颗这样的头骨时，关于“脑手术”的说法才得到证实及认可。

这个发现证实了新石器时代的人类确实会进行颅骨穿孔术。头骨上留有的手术痕迹表明，手术是针对活人的，而不是死后的宗教行为，甚至其中一些人在经过多次外科颅骨手术后仍然活着。但是这种手术的目的是什么呢？有些人相信，通过在头皮与头盖骨上钻孔的方式，可以释放颅内过大的压力；有些神秘主义者认为，在头盖骨上打洞，可以提升感应能力；而在有些宗教信仰者眼里，头盖骨上的洞也许是为了给邪恶的灵魂打开一个离开脑子的通路……这些说法反映了在当时的欧洲，关于大脑的手术已经被一些外科医生用来治疗精神类疾病，或者作为一种躲避恶灵的手段。不过，古印加的外科医生做这种手术的真实意图是什么，我们至今不得而知。

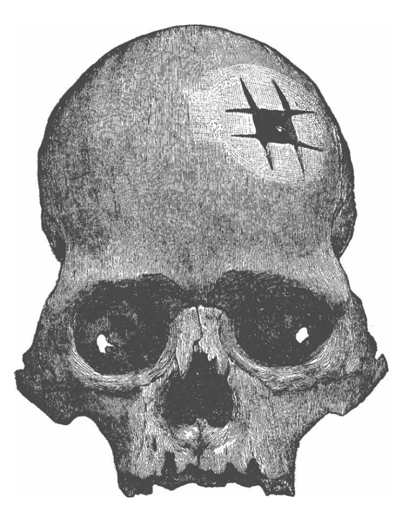


图1-1　被陈列在美国自然历史博物馆的穿刺头骨

中国古代也有诸多学者试图探究人脑与心理活动的关系。其中，早在战国时期的《黄帝内经》便已涉及脑的解剖构造。《灵枢·海论》说："脑为髓之海，其输上在于其盖，下在风府。"不但指出脑是髓汇集而成，而且认为脑与脊髓相连，与全身的髓都有密切的关系，故《素问·五脏生成篇》说："诸髓者，皆属于脑。"因此，脑也被称为"髓海"，这就是"脑髓说"的萌芽。后世的"脑髓说"认为，大脑是精髓和神明汇集发出之处，又称"元神之府"。《灵枢·海论》中还说："髓海有余，则轻劲多力，自过其度。髓海不足，则脑转耳鸣，胫酸眩冒，目无所见，懈怠安卧。"这说明了脑对人体机能有着直接的影响。

然而，不论是西方还是中国，人们起初都将心脏视为记忆和思考的器官。《礼记》中记载："心不在焉，视而不见，听而不闻，食而不知其味。"古埃及会在人死后将其制作成木乃伊，目的是希望灵魂能够找回躯体而顺利复活，尽管死者尸身可以被保存得十分完好，但他们的大脑却在制作木乃伊的过程中被从鼻腔中取出丢掉。

这种"心脏是灵魂居所"的想法直到古希腊时代才受到强有力的挑战，并随着对人体的进一步探索和科技的发展逐渐改变。直至1795年，人们才确定人类的思维来自大脑。这对我们现代人来说是常识，在那个时期却有着跨时代的意义。在技术水平有限的年代，人类对大

脑的探索总是磕磕绊绊，下面，就让我们坐上时光机，一起见证脑科学的前世今生吧！

## 神经科学的诞生

对于大脑的研究起始于它的结构层面，如果没有结构知识作为基础，对大脑功能的探索就如同空中楼阁、沙上建塔。在对大脑结构的探索过程中，神经科学应运而生了。

### 神经科学的萌芽

时光机的第一站是公元前4世纪的古希腊。正如前文所说，在这个时代，“心脏是灵魂居所”的想法受到了强有力的挑战。西方医学奠基人、“医学之父”——希波克拉底（Hippocrates）（公元前460年—前379年）通过对“结构-功能相关性”的思考以及解剖观察，得出了“思维、情感、智慧皆来自于大脑，大脑参与对环境的感知”的结论。但这一观点并未得到普遍的认可，例如，著名的古希腊哲学家亚里士多德（Aristotle）就固执地相信“心脏是智慧之源”。他认为大脑仅是一个散热器，被“火热的心”沸腾的血液在这里得到冷却，并以此解释了人体恒定且合适的体温。

时光机的第二站是古罗马时代。古罗马医学史上最重要的一位人物——盖伦（Galen）（130年—200年）接受了希波克拉底关于脑功能的观点。同时，根据对大量动物细致地解剖（特别是羊脑），他提出将大脑分成3个腔室，分别承担想象、推理和记忆这3个心理过程。这些腔室被称为脑室（类似于心脏的心室），大脑通过这3个脑室泵出液体，来控制身体不同的活动。在盖伦看来，这一发现极好地吻合了当时流行的理论：神经是一种类似于血管的中空管道，机体的功能有赖于4种重要液体的平衡，液体通过神经管道流入或流出脑室，使大脑得以执行不同的功能。

时光机的第三站是文艺复兴时期。盖伦有关于大脑的观点延续了将近1300年，直到文艺复兴时期，法国近代解剖学创始人安德烈·维

萨里（Andreas Vesalius）（1514年—1564年）出版了第一部真正记载神经科学的医学巨著——《人体的构造》。至此，医学界对人体的认知，终于从由动物推论变成了从人体本身出发，神经解剖学就此建立，人们对大脑结构的认识也逐渐精细化。

虽然维萨里在《人体的构造》中进一步补充了许多脑结构方面的细节知识，但是却没有挑战脑功能的脑室观点。相反，由于17世纪早期法国人开始使用以水为动力控制的机械装置，脑功能的脑室观点又得到了进一步的强化。这些机械装置支持了“以类似于机械运行的方式行使其功能”的观点：液体从脑室中被压出，经过“神经管道”到达人体各处，从而激发肢体的运动。法国数学家和哲学家勒内·笛卡儿（René Descartes）便是这一观点的主要提倡者。

不过，尽管他认为这一理论可以解释其他动物的脑和行为，但用该理论去解释人类所有的行为却是一件不可思议的事情，因为与其他动物不同，人类拥有智慧和一颗上帝赐予的心灵。因此，笛卡儿提出，尽管大脑是控制身体行动的器官，但人类所特有的“心灵”则独立于大脑之外，人类的灵魂、思想，都跻身于此。与此同时，大脑与心灵通过大脑内的一个叫松果体的结构（实际上是脑内的一个分泌各类激素的结构）进行交流。他的这种说法，无论在哲学界，还是在神经科学界，都影响颇深。直至今日，仍有人相信“心灵”与脑是彼此分离的。但是，正如我们将在本书后续关于脑的认知功能中介绍的那样，现代神经认知科学并不支持这种说法。

接下来我们来到时光机的第四站——17—18世纪。一些科学家挣脱了盖伦的脑室论这一传统观念的束缚，对脑结构进行了更加深入的研究。他们观察到脑组织可被分为两部分：灰质和白质，且正确地提出白质包含纤维，这些纤维起到向灰质传递信息的作用。

到18世纪末，神经系统已经可以被完整地剥离出来，它的大体解剖也因此获得了更为细致的描述。神经解剖学史上的一个重大突破是在脑表面观察到广泛存在的一些凸起和凹槽，它们被分别称为脑回和脑沟（在第2章中会详细介绍）。这一结构使大脑可以以脑叶的形式进行划分，奠定了“不同脑功能定位于不同的脑回”的理论基础，为脑功能定位研究开创了新时代。

## 颅相学的兴与衰

我们已经见证了脑功能定位研究新时代的开启，现在我们将重点介绍一个曾风靡欧美的脑功能定位假说——颅相学。

颅相学与我国古代的面相学类似，是一门通过研究人体颅骨外部形状来判断一个人的性格和命运的学说。1796年，德国解剖学家弗兰茨·约瑟夫·加尔（Franz Joseph Gall）（图1-2）首次提出了颅相学的概念。在他看来，头骨和大脑的形状是紧密对应的，某个特定脑区的大小直接决定了头骨的形状，因此，如果对头骨的凹凸形状进行分析，就可以了解到每个人的性格和能力。比如隆起的头顶代表着智慧，宽阔的前额说明想象力丰富，而大头则意味着聪明绝顶。

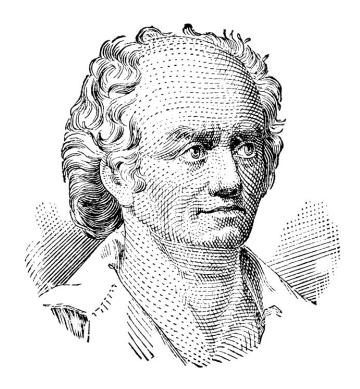


图1-2 弗兰茨·约瑟夫·加尔画像

加尔从小就对面部和颅部特征非常感兴趣。高中时代，他发现几位记忆力出众的同学，眼睛非常突出，据此，他推断位于眼睛后方的脑区应该与人的语言和记忆有所关联。之后的许多年里，加尔通过这样类似的观察归纳，总结出了27个功能区域，如图1-3所示。19世纪初，加尔开始发表有关颅相学理论的医学文献。他的研究结果在推动人类大脑研究的同时，也推动了人类对自身以及与其他动物之间差异的认知。

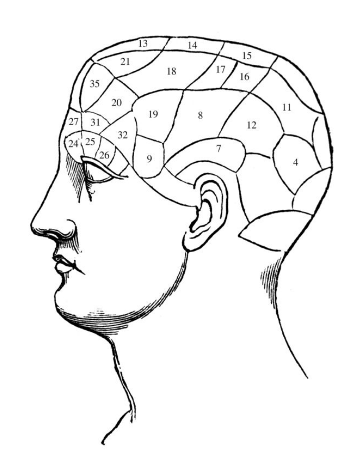



图1-3　颅相学示意图

19世纪20—40年代，颅相学正处于发展的鼎盛时期。在学术界，颅相学获得了一些杰出的科学家，甚至医学界的领军人物的认可；在政治界，英国女王亚历山德丽娜·维多利亚（Alexandrina Victoria）以及美国总统约翰·亚当斯（John Adams）都欣然接受颅相学大师的诊断；而在普通民众的生活中，颅相学诊所在欧美大街小巷四处开花，不仅谈婚论嫁需要去看颅相，而且找工作时，许多雇主也都要求求职者提供一份由当地的颅相学家出具的性格证明，以确保未来的雇员诚实、勤奋。头骨上的凸起提供了一个判断人才和能力的指标，这一信念尤其被用于教育和刑事改革。头部的形状与大小俨然成为欧美民众沉迷讨论的话题。

不过，即使再繁华的高楼也会一夕崩塌。随着对医学、生物学领域的研究越来越深入，各界对加尔颅相学的质疑声也越来越大。

法国的神经生理学家皮埃尔·让·玛丽·弗卢龙（Pierre Jean Marie Flourens）是颅相学理论最大的反对者之一。他对鸽子进行脑部切除手术时，发现不论什么位置的小部分损毁，鸽子仍然能吃能睡，看上去并无大碍；而当鸽子脑部被切除的面积越来越大时，鸽子才开始逐渐出现异常。因此弗卢龙认为大脑其实是作为一个整体运行的，每个区域都均等地参与了所有脑功能，无法单独通过某个区域独立运作。这个说法显然与颅相学中“不同位置的头颅区域代表着不同能力”相悖。此外，弗卢龙通过解剖还得出大脑和头骨形状并不是一一对应的

结论。从此，颅相学开始由“众人追捧”逐渐走向“众人追喷”。

颅相学虽然衰落了，但是其关于脑功能定位的见解却依然影响着后人，人们关于大脑功能的“定位说”与“整体说”也一直争论不休。最终，法国神经科医生皮埃尔·保尔·布罗卡（Pierre Paul Broca）使科学的天平稳稳地偏向大脑功能定位说的一侧。布罗卡曾经遇到过这样一个病人，他能够理解别人的言语，自己却无法说话。在这个病人死后，布罗卡仔细地研究了他的大脑，结果在其左额叶上发现了损伤。根据这一病例以及其他几个类似的病例，布罗卡认为大脑的这一区域具体负责语言的形成，并将其命名为broca区（布罗卡区），如图1-4所示。

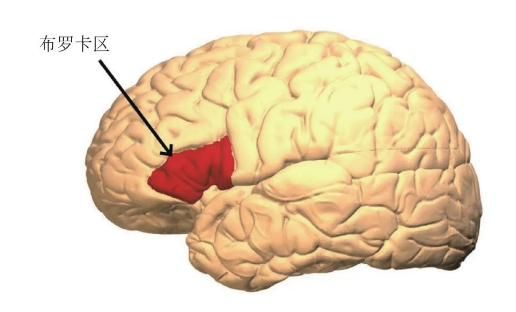



图1-4　布罗卡区示意图

从积极的角度看，颅相学的确是第一个提出“大脑功能及空间分布关系”这一观点的学说，后来布罗卡发现大脑语言中枢，也在一定基础上保留了颅相学的观点。但由于缺乏现代神经科学的工具，当时的科学家只能利用观察来进行小范围的研究，很有局限性。这些细节上的错误，导致颅相学走向了荒谬可笑的方向，最终被时代淘汰。利用磁共振等现代技术，今天的神经学家可以重新审视和探索大脑的不同区域以及它们与不同功能和心理特征之间的联系，这也是当下脑科学研究的热点方向。关于脑功能定位的现代研究方法，我们在第3章进行详细介绍。

## 神经元的发现

正如曾经风靡一时的颅相学最终走向衰落的结局告诉我们的那

样：技术的不足会限制我们对事物的观察，而技术的突破一般都可以帮助各种科学理论更进一步发展和完善。到19世纪中期之前，人们对大脑的认知还停留在形状、大小这类宏观的层面，对大脑的构成并不了解。实验仪器精准度的限制是一个很大的原因。当高精度的显微镜被发明之后，科学家们终于能看清楚神经系统了。随着生物细胞理论的发展，人们认识到，大脑组织也是由细胞构成的。

起初，就算是有高精度的显微镜，大脑组织在显微镜下也只是一堆不太能被区分的颗粒状的组织，所以在当时仍有很多人反对大脑是由细胞构成的这一观点。后来意大利解剖学家卡米洛·高尔基（Gamillo Golgi）（就是发现细胞中高尔基体的那个高尔基）发明了一种银染色法（高尔基染色法），来标记脑神经细胞。西班牙人圣地亚哥·拉蒙·卡哈尔（Santiago Ramóny Cajal）使用高尔基染色法发现神经元是分立的个体。他不仅第一次鉴别出了神经元的单一性，而且还发现神经元内的电传导是单向的，只能从树突传到轴突。

在发现神经传导路线的同时，卡哈尔也提出了神经细胞是通过突触结构来划分的，即大脑也是通过大量独立细胞所组成的组织，形成了后来著名的“神经元学说”，他本人也被称为“现代神经科学之父”。至此，现代神经科学终于诞生了！

## 走近认知科学

我们乘坐时光机见证了神经科学的诞生。随着20世纪神经科学的不断发展，脑功能定位主义者对他们的观点进行了一定的取舍。他们发现，尽管特定的大脑区域负责某项独立的功能，但这些区域组成的网络以及它们之间的相互作用才是产生人类表现出整体、综合行为的原因。

法国生物学家克洛德·贝尔纳（Claude Bernard）曾说过：“如果有可能将身体的所有部分分解开，将它们独立出来以研究它们的结构、形式和连接，那就和生命不同了……如果一个人只分别研究一种机制的各个部分，那么他就不可能知道它是如何运作的。”

就这样，科学家们开始相信，关于神经元和大脑结构的认识，必须放在整体的关系中被理解，即当这些部分连接到一起时产生的作用。因此，神经科学的基本方法并不能全面地分析大脑，因为大脑是一个活的、动态的系统。接下来，我们就从另一个科学的角度——认知科学，来认识脑科学的历史。

## 心理学的故事

尽管脱胎于医学的神经科学在大脑研究的早期阶段引领潮流，但心理学家对心智的研究早已经通过测量行为进行着，而这就是认知科学的前身。

在实验心理学诞生之前，对心智的探讨一直是哲学家的领域，他们对知识的本质以及人类如何认识事物充满好奇。哲学界有两大主要观点：理性主义和经验主义。理性主义兴起于启蒙运动时期，在知识分子和科学家中，理性主义取代了宗教，成为思考世界的唯一方式。理性主义者从自然科学中吸取辩证法发展的观点，用思辨的方式来表达进步的要求。相反，经验主义者认为，所有的知识来自于感觉经验，直接的感觉经验可以产生简单的思想和观念。当各种简单的想法相互作用、相互连接，复杂的想法和观念就产生了。

随着历史的推进，一方面，从古希腊到19世纪中叶近2000多年的哲学发展已经为心理学的独立酝酿了必要的条件；另一方面，19世纪西方科学的发展已经有了长足的进步，当时的生理解剖学、物理学等许多自然科学获得了巨大的进展，它们确立了科学的权威地位，同时也为心理科学的独立创造了条件。如著名的心理物理学家费希纳（Fechner）和韦伯（Weber）通过实验，将事物的物理性质（光和声音），同它们给观察者造成的心理体验联系起来。这些实验使得一些心理学家意识到，要想使心理学从哲学中脱离出来成为一门独立的学科，就必须把这些科学方法引入心理学的研究，而这也是使心理学成为科学的最直接的前提条件。1879年，威廉·冯特（Wilhelm Wundt）在莱比锡大学建立了第一个心理学实验室，这意味着现代实验心理学的开始，也意味着科学心理学的确立。

然而，在之后的几十年中，实验心理学却开始被行为主义统治。

大家一定知道著名的巴甫洛夫实验：当不断地把铃声和喂食匹配在一起，狗会逐渐对铃声流口水。这似乎在预示着物理刺激和心理学习过程可以被精心地控制并有效地测量，就像马戏团训练动物表演那样，行为实际上是重复地物理刺激的训练结果。



巴甫洛夫的狗

行为主义之父——约翰·华生（John Waston）将巴甫洛夫的条件反射学说作为学习的理论基础。他认为学习就是以一种刺激替代另一种刺激建立条件反射的过程，并宣称在环境完全可控的情况下，他可以把一个孩子塑造成任何样子。在华生看来，人类出生时只有几个反射行为（如打喷嚏、膝跳反射）和情绪反应（如惧、爱、怒等），其他所有行为都是通过条件反射建立新刺激-反应联结而后天习得的。为了让大家接受这个观点，他设计了一个臭名昭著的、令人心碎的实验——小阿尔伯特实验（little Albert experiment），试图证明情绪可以经由条件作用而产生，不用考虑任何内部的力量。

实验之前，华生这样描述他的受试者阿尔伯特："一个重9.5千克，11个月大的婴儿……他健康、温和，是个妙极了的好孩子。在与他相处的几个月中，我们从来没有看见他哭过，直到我们做了实验之后……"

在阿尔伯特9个月大时，实验者向他呈现大白鼠、兔子、狗、棉毛织物等东西，来观察他对这些特定刺激的反应。结果发现阿尔伯特不但没有表现出任何恐惧情绪，反而十分感兴趣，时不时地抚摸这些物品，此时，这些物品还属于中性刺激。然后，实验者开始测试阿尔伯特对巨大噪声的恐惧反应。他们在阿尔伯特身后用锤子击打钢棒，制造出响亮并吓人的噪声。可想而知，阿尔伯特被吓坏了，他在巨大噪声的刺激下爆发大哭。

正式实验在阿尔伯特11个月大时开始。当阿尔伯特伸手去触摸一只大白鼠时，实验者在一旁用锤子击打钢棒。钢棒发出巨响，阿尔伯特被吓得猛地跳了起来，跌倒在床上。此后，每当他要伸手触摸大白鼠时，实验者便敲击钢棍，将他吓得猛然跳起然后跌倒，继而大哭。最后，“只要白鼠一出现，婴儿就开始哭。他几乎立刻就……开始爬得飞快，以至于在他爬到桌子边缘时差点儿没能拉住他”。就这样，阿尔伯特对噪声的自然反应变成了对白鼠的条件反射。

基于这一发现，华生和他的助手想知道阿尔伯特对白鼠的恐惧是否会转移到其他毛茸茸的动物身上，于是他们开始将兔子拿给他。结果阿尔伯特还是一边哭一边爬走了。甚至从此对于狗、白色毛皮大衣、绒毛娃娃、棉花等毛茸茸的东西，阿尔伯特都产生了深深的恐惧——这也是这个实验臭名昭著的原因。

显然华生的实验是残忍的、不符合伦理的，但是他的行为主义心理学却一时成为心理学主流，影响美国心理学长达30年之久，直到20世纪50年代才真正结束。

实际上，行为主义心理学乍看上去似乎有些道理，但仔细想想便能发现许多漏洞。如著名的托尔曼（Tolman）老鼠迷宫实验，便是反击行为主义心理学的一大有力证明。如图1-5所示，迷宫有1个出发点、1个食物箱和3条长度不等的从出发点到达食物箱的通道（分别为通道1、通道2、通道3）。实验开始时，先让小白鼠在迷宫内自由地探索，一段时间后，检验它们的学习结果。

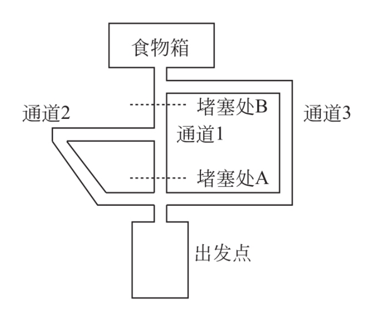



图1-5　小白鼠学习方位的迷宫

当3条通道都畅通时，小白鼠会选择距离最短的第一条通道，也

就是说，在一般情况下，小白鼠往往选择较短的途径。而当爱德华·切斯·托尔曼（Eduard Chace Tolman）对各通道做一些处理后，例如，在A处将通道1堵塞，这时发现小白鼠选择通道2跑到食物箱；当在B处堵塞通道1时，小白鼠并不像以前形成的习惯那样，先选择通道2再选择通道3，而是避开通道2，马上选择通道3。即小白鼠能“顿悟或意识到”堵塞B点会将通道1与通道2同时关闭起来，就像它们的头脑中存在迷宫地图一样。

根据实验结果，托尔曼认为小白鼠走迷宫，学习的并不是左转或右转的序列，而是在它的脑中形成一种认知地图，如果一条熟悉的路被堵塞，小白鼠就会根据认知地图所展现的空间关系选择另一条路线到达目标。而“大脑可以快速产生没有被训练过的行为”这一现象，显然无法被行为主义解释。

除此之外，我们都知道，语言具有复杂性与多样性，人们可以将同一个意思用很多个不同的语句甚至不同的语言表达。因此，心理学家们逐渐认识到打开大脑这个黑盒子的重要性。老鼠头脑里的迷宫地图、抽象的行为目标、语言学等新挑战开始让下一代心理学家重新思考研究的框架。

## 认知心理学

如果说行为主义时代信奉的“心理学是研究行为的科学”是片面的，那么到底什么才是心理学呢？

乔治·米勒（George Miller）为一时有些茫然的心理学界指点了迷津。1960年，他与另一认知心理学家杰罗姆·布鲁纳（Jerome Seymour Bruner），联合成立哈佛大学认知研究中心，该中心的命名即带有向行为主义心理学挑战的意味。不过米勒本人并不赞同人们将他们的心理学思想解释为认知革命。他认为认知心理学的兴起并非完全创新，只能说是旧思想的复苏。认知心理学把以往被行为主义心理学排挤到后台的人的认知过程重新拉回到心理学研究的前台，重视对注意、知觉、表象、记忆、思维和语言等高级心理过程的研究，从而使心理学恢复了原来研究内在心理活动的本来面貌。自此，心理学便

从原先的“心理学是研究行为的科学”，改变为“心理学是研究行为与心理历程的科学”。

米勒很清晰地记得，自己当初下定决心放弃行为主义心理学而转向认知主义的那一天是1956年9月11日，麻省理工学院举办第二届信息理论研讨会期间。对很多学科来说，那一年是个丰收年。

例如，当时的计算机科学领域就发展十分迅速，艾伦·纽厄尔（Allen Newell）和希尔伯特·西蒙（Herbert Simon）成功提出了“第一代信息加工语言”，并开发了最早的启发式程序“逻辑理论家”和“通用问题求解器”——一个强大的、可以模拟逻辑定理证明过程的程序。处理计算机信息功能的改变，对尚处于萌芽时期的认知心理学产生了重大影响。

尽管在这之前，心理学家已经将信息处理的历程大致区分为感官记忆（2秒以下）、短时记忆（15秒以下）和长时记忆，但短时记忆的性质及其重要性，则是在米勒于1956年发表研究报告《神奇的数字7±2：我们信息加工能力的局限》之后才被确定的。

米勒受到计算机处理信息方式的启发，提出了信息编码的概念。他认为编码最简单的方式是将输入信息归类，然后加以命名，最后储存的是这个命名而非输入信息本身。编码是一个主动的转换过程，对经验并非严格的匹配，因此编码以及解码往往会导致错误发生。他的研究有两点要义：

第一，在不得重复练习的情形下（如看电视字幕），短时记忆中一般人平均只能记下7个项目（如7位数字、7个地名），因此，从电话簿上查到一串电话号码后，往往在要拨号时会不复记忆。

第二，短时记忆的量虽然不能增加，但却有可能根据所记忆事物的性质经由心理运作使之扩大。例如，2471530122022是一串13位的数字，远超过“7”这个数量限制，但如果经心理运作将之意义化：24小时（一天）、7天（一星期）、15天（半个月）、30天（一个月）、12个月（一年）、2022（年份）——是不是就变得容易记忆了？米勒称此种意义单位为组块，人们学习英文时由字母组成单词，由单词组成短语，由短语组成长句……这些都是将零碎信息经心理运

作变成多个组块之后记下来的。

米勒对短时记忆上的研究成就，为新兴的认知心理学提供了理论的依据。自此之后，短时记忆成为现代认知心理学中热点主题。米勒为以信息加工理论研究记忆开创了道路，而他的学术成果甚至成为谷歌等公司搜索技术的研究基础。

现在，认知心理学已经成为心理学、认知科学甚至脑科学研究中的重要组成部分。不仅如此，很多互联网公司，特别是游戏公司，在开发新应用和新产品的时候，也会从认知心理学的研究成果中汲取灵感。另外，经济学领域也开始被认知心理学渗透，畅销书《思考快与慢》的作者丹尼尔·卡内曼（Daniel Kahneman）便是首位获得诺贝尔经济学奖的认知心理学家。

## 认知科学的今生

就这样，以乔治·米勒为代表的一批心理学家将计算的思想带入对大脑的研究当中，以信息处理论为基础的认知科学便以此为沃土，逐渐兴起了。1977年，《认知科学》的创刊与1979年认知科学学会的成立，标志着认知科学开始渐渐走近大众的视野。

认知的英文是cognition，它来源于拉丁语，是“了解、学习”的意思，因此，认知过程讲述的就是我们如何学习和了解外部世界，如何处理信息的过程。这个过程涵盖了注意力、记忆、觉察、语言、元认知等更具体的内容。而由其发展历史可见，认知科学是研究心智和智能的交叉学科，是现代心理学、人工智能、神经科学、语言学、人类学乃至自然哲学等学科交叉发展的结果，也是脑科学的一大重要研究方向。

认知科学的研究内容主要包括：①以知觉表达、学习和记忆过程中的信息处理、思维、语言模型和基于环境的认知为突破口，在认知的计算理论与科学实验方法与策略等方向实现原始创新；②探讨创新学习机制，建立脑功能成像数据库，提出新的机器学习和方法。由于其涉及学科之广、研究前景之大，认知科学的发展得到了国际科技界，尤其是发达国家政府的高度重视和大规模支持。

21世纪初，美国国家科学基金会和美国商务部共同资助了一个雄心勃勃的计划——“提高人类素质的聚合技术”，他们将纳米技术、生物技术、信息技术和认知科学看作21世纪四大前沿科技，并将认知科学视为最优先发展的领域。美国海军支持认知科学的规划——“认知科学基础规划”，已有30多年的历史。其基本目标包括5个方面：①确定人类的认知构造；②提供知识和技能的准确认知结构特性；③发展复杂学习的理论，解释获得知识结构和复杂认知处理的过程；④提供教导性理论以刻画如何帮助和优化学习过程；⑤利用人类行为的计算模型，提供建立有效的人-系统交互作用的认知工程的科学基础。

世界一流大学都已经开展了认知科学的研究，并在各自的研究范围取得了丰硕的成果。中国对认知科学的研究也很重视，目前已建立了若干个与认知科学和智能信息处理密切相关的国家重点实验室和一批省部级重点实验室，形成了包括若干个知名院士和一批优秀中青年科学家在内的研究队伍，相关实验室的软硬件装备已接近或达到世界先进水平。中国于2001年正式成为“人类脑计划”的会员国之一，并已经与认知科学界建立了广泛和实质性的国际合作与交流。

总之，探究人类心智始终是科学家孜孜不倦的追求。毕竟之前很长一段时间里，人类对心智的探索只停留在哲学、心理、解剖学等层面，随着计算机科学等新兴学科的建立，让科学家看到了从学科融合的角度切入去研究人类心智，认知科学也就应运而生了。

## 脑科学的未来

上文我们提到，以乔治·米勒为代表的一批心理学家将信息处理论带入对大脑的研究当中。认知革命之后，“人类认知系统（大脑）是一个信息加工系统”这一观点已被许多科学家认可，而对于这个系统的研究自然也就有不同的角度，即结构和功能。

我们以一个大家更熟悉的信息加工系统——计算机为例，假设有一台时空机器把现代的计算机传送到100年前——100年前的人并不知道这台计算机的构造和原理——但是他们肯定觉得这东西特别有意思，于是就会有一群科学家坐下来研究这台神奇的东西。

首先，会有一群人坐下来拆机器，他们拆开计算机的外壳，看到里面的结构和部件：中央处理器（central processing unit, CPU）、随机存取存储器（random access memory, RAM）、数据线路等。他们想知道这台神奇的机器是怎么运行的，需要对这台机器的结构组成有所了解。我们称这些人为硬件科学家。硬件科学家可能会通过不同的实验，比如拆掉某个硬件单元来研究这个硬件单元对整个机器工作的影响，或者研究每一个部件的构造，看看他们的工作特点等。

不过，硬件科学家做的事情，并不能告诉我们计算机里面的操作系统是怎么被编写出来的。所以，一群思路截然不同的科学家也加入进来研究计算机，他们决定暂时不考虑硬件的工作方式，直接研究桌面上的一个个软件，我们称之为：软件科学家。软件科学家并不关心他们所看到的桌面操作系统是怎么通过物理元件实现的，他们更关心的是这个操作系统具有什么功能，能够做什么事情，具有什么性能。

硬件科学家和软件科学家都在研究这台被我们传送过去的计算机，研究它是怎样工作的，但是很显然，硬件科学家和软件科学家在做完全不同的事情，有着完全不同的研究方向。

现在我们回到脑科学上来。同样地，我们也有一批硬件科学家和软件科学家正在研究我们的大脑。简单地说，神经科学更像是“硬件学派”，而认知科学更像是“软件学派”。前者关心的是“大脑”这个物理系统是怎么样进行信息加工，从而执行人类当前行为的；后者关心的是“认知系统”需要执行什么样的运算才能产生人类当前的行为。

在前文，我们已经从脑科学的“硬件学派”和“软件学派”介绍了它的发展历史。接下来我们要介绍的是脑科学的研究现状以及未来。

## 现在：脑计划更懂你

脑是人体最复杂的器官，负责对人体一切行为、思维、决策和感觉的调控。然而，目前人类对大脑的了解尚处于初级阶段。只有更好地了解大脑的构造和功能，才能对脑部的疾病做出更完善的诊断和治疗。此外，随着计算机技术的发展，人工智能被推到时代发展的风口，通过借鉴大脑神经网络，可以更好地促进人工智能的完善。

正是基于这些来自医疗、科研和技术的需求，人类脑计划应运而生。自2013年起，美国、欧盟、日本相继启动了各自的大型脑科学计划，全球参与脑计划的国家数量不断扩充壮大，它不仅仅是科技发展的信号，更代表了全球化科研资源的整合（表1-1）。

表1-1　脑计划简介

| 发起方 | 脑计划名称 | 启动时间 | 研究重点 |
|---|---|---|---|
| 美国 | 创新性神经技术大脑研究计划<br>BRAIN, Brain Research through Advancing Innovative Neurotechnologies | 2013 年 | 旨在推动创新技术的开发与应用，研究大脑动态功能及工作机制，开发治疗脑部疾病新方法 |
| 欧盟 | 人类脑计划<br>HBP, Human Brain Project | 2013 年 | 旨在利用超级计算机技术模拟大脑功能，从而实现人工智能 |
| 日本 | 疾病研究综合神经技术脑图绘制<br>Brain/MINDS, Brain Mapping by Integrated Neurotechnologies for Disease Studies | 2014 年 | 旨在通过融合灵长类模式动物多种神经技术研究，以研究人类神经生理机制，并建立狨猴脑发育及疾病发生的动物模型 |
| 澳大利亚 | 澳大利亚脑计划<br>Australian Brain Initiative | 2016 年 | 旨在揭示脑异常机制、编码神经环路与脑网络认知功能，解决人类健康、教育问题，并通过促进工业合作者和脑研究的结合研发新的医疗产品 |
| 中国 | 脑科学与类脑科学研究<br>Brain Science and Brain-Like Intelligence Technology<br>简称为“中国脑计划”<br>China Brain Project | 2016 年 | 形成“一体两翼”布局，以研究脑认知的神经原理为“主体”，以研发脑重大疾病诊断新手段和脑机智能新技术为“两翼” |
| 韩国 | 韩国脑计划<br>Korea Brain Initiative | 2016 年 | 旨在破译大脑的功能和机制，调节作为决策基础的大脑功能的整合和控制机制 |

中国科学家在由中国科学技术部和自然科学基金委员会组织举办的许多战略会议上进行了讨论，最终达成了一个共识，即神经科学的

一个普遍目标——理解人类认知的神经基础——应该成为“中国脑计划”的核心。由于主要脑疾病造成的社会压力逐渐上升，所以现在迫切需要一种预防、诊断和治疗脑疾病的新方法。此外，在大数据的新时代，受大脑启发而获得的计算方法和系统对于实现更强的人工智能和更好地利用越来越多的信息至关重要。正是由于对这些问题的考虑，中国脑计划（图1-6）项目提出了“一体两翼”战略。即以研究脑认知的神经原理为“主体”（其中又以绘制脑功能联结图谱为重点），以研发脑重大疾病诊治新手段和脑机智能新技术为“两翼”。

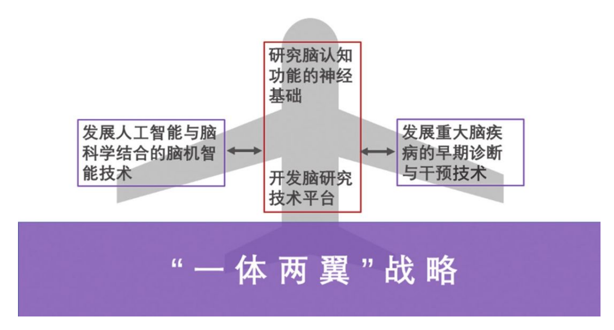



图1-6　中国脑计划

“脑科学与类脑科学研究”，简称为“中国脑计划”，于“十三五”期间正式启动。《中华人民共和国国民经济和社会发展第十三个五年规划纲要》提出强化脑与认知等基础前沿科学研究，将脑科学与类脑研究纳入“科技创新2030——重大项目”。在2021年，科技部网站发布通知称，科技创新2030“脑科学与类脑科学研究”重大项目年度申报涉及59个研究领域和方向，国家拨款经费预计超过31.48亿元。《中华人民共和国国民经济和社会发展第十四个五年规划和2035年远景目标纲要》提出，瞄准脑科学等前沿领域，实施一批具有前瞻性、战略性的国家重大科技项目，攻关脑认知原理解析，脑介观神经联接图谱绘制，脑重大疾病机理与干预研究，儿童青少年脑智发育，以及类脑计算与脑机融合技术研发。从“十三五”规划到“十四五”规划，脑科学都被列为重点前沿科技项目。不仅如此，《“健康中国2030”规划纲要》也提出启动实施脑科学与类脑研究等重大科技项目和重大工程。脑科学作为我国一个相对较新的研究学科，目前正处于积极发展阶段，此

前世界上正在实施的脑计划的经验和教训对我国进行脑计划也有所帮助。

面对脑科学这一自然科学的“最后疆域”，需要的是整个国际科研界的成果共享和通力合作。相信在各个参与方各具特色的脑计划共同协作下，人类对脑和疾病的认知将不断深入，并从中寻找到更高的应用价值。

## 未来，可能比科幻更科幻

现在，让我们一起畅想一下脑科学的未来。实际上，关于未来，人们总是有各种各样的畅想，尤其是关于人脑的未来，已经有许多影视作品和科幻小说对其进行了描述。

## 掌控梦境

梦境向来是一个神秘又引人好奇的领域，千百年来，关于梦的传说也不尽其数。一些科学家曾在梦里获得灵感。药理学家奥托·勒维（Otto Loewi）在梦中获得了神经递质能够促进信息通过突触的想法，而这成为了神经科学的基础。无独有偶，德国著名有机化学家奥格斯特·凯库勒（August Kekulé）在1865年做了一个关于苯的梦。他梦见苯的碳原子构成一个链条，首尾相接，形成环状，就像一只咬着自己尾巴的蛇。基于这个梦，他提出了苯分子的物质结构，即大名鼎鼎的苯环。



凯库勒的衔尾蛇

一些影视作品中也常常出现梦的题材，在经典电影《盗梦空间》中，莱昂纳多·迪卡普里奥（Leonardo Dicaprio）饰演的男主角可以从梦这个最不可能的地方盗取秘密。他的团队可以利用一种新的发明进入人们的梦境，从人的潜意识中盗取机密，并重塑他人的梦境。

尽管梦一直困扰和迷惑着我们，但科学家似乎已经抓住了解梦的蛛丝马迹。事实上，科学家现在所做的一些事曾经被认为是不可能的：他们可以用磁共振成像技术拍下梦的模糊图像和影像（关于这项技术的原理我们会在第3章中进行介绍）。也许有一天，你可以通过观看自己梦的视频了解自己的潜意识；又或许，经过适当的训练，你可以有意识地控制自己梦。甚至在更远的未来，通过计算机对两个正在做梦的大脑进行磁共振成像扫描，并将其中一个人的扫描结果解码为视频影像，传输到另一个人大脑的知觉区域，这样，二者的梦就可以进行合并，或许就能实现像《盗梦空间》中的角色那样潜入别人的梦境。

## 意念互联

“脑机接口”这一话题在近年成为热点。这种技术可以将大脑中的神经元信号转换为能够在现实世界中移动物体的具有实际意义的指令，为残障人士重新恢复一部分人体机能提供了可能。著名宇宙学家斯蒂芬·威廉·霍金（Stephen William Hawking）就安装了一个类似脑机接口的设备。这一设备就像一台脑电图传感器一样，能够将霍金的思维和计算机连接在一起，这样一来，他就能保持自己同外部世界的联系了。除了可以用于改善病人生活，这一技术的另一种用途是把计算机与任意设备连接起来，并实现脑控，如烤面包机、咖啡机、空调、电灯开关等。有了这种技术，我们就可以实现坐在家中，仅仅动动脑子便自由切换电视频道、开关灯以及烹饪料理了。关于脑机接口的更多知识，我们将会在本书的第4章进行介绍。虽然目前这种技术还处于起步水平，但随着科技发展，也许“意念互联”就在不远的明天。

## 植入记忆

在《黑客帝国》中，主人公尼奥（Neo）可以通过脖颈上植入的电极，即时将武术技能下载到大脑中，仅仅几秒钟，他便成了跆拳道大师，轻而易举地打倒了追杀他的人。你一定也在考试前有过这样的

幻想：如果能够下载记忆，我就不用复习了！

这样的情节看似天方夜谭，但也许真的可能成为现实。2013年麻省理工学院的一个课题组在研究阿尔茨海默病时发现，他们不仅能够实现在老鼠的大脑中植入普通记忆，还可以实现植入虚假记忆。这些科学家使用了一种叫作光遗传的技术，通过对特定的神经元进行照射，从而使其激活。利用这种技术，科学家能够识别出对特定记忆而言是哪些特定的神经元在起作用。例如，一只老鼠进入房间，然后被电击。科学家可以分离出承受这个痛苦记忆的神经元，通过分析海马体把它记录下来并与光纤维连接。然后，把这只老鼠放进一间完全不同且绝对安全的房间里，打开光源照射光纤维，老鼠便会在这个新房间中产生虚假的电击记忆，并做出恐惧的表现。

不过，遗憾的是，技术的发展会在一定程度上限制人类的想象力，就像100多年前的人没有互联网的概念，更无法想象什么是“万物互联”。不过，我相信脑科学的未来，也许比科幻更加科幻。

## 小结

在这一章中，我们乘坐时光机回顾了神经科学与认知科学的诞生与发展历史。作为脑科学的“硬件学派”和“软件学派”，神经科学与认知科学的历史正是脑科学的历史。除此之外，我们还简单地对脑科学的未来发展进行了一些展望，正是这些看起来既神秘又酷炫的未来，吸引着无数科学家孜孜不倦的研究。如果你对大脑也同样感兴趣，就请继续读下去，希望对你认识脑科学有所帮助。

大脑犹如我们人体的司令部，支配着我们的思想、行动、情绪，乃至潜能，其重要性不言而喻，而脑科学却能让我们去揭开这层神秘的面纱。脑科学为什么很酷？Because it is the only case in the world when an operating system is attempting to study itself（因为世界上只有在这个学科中，一个操作系统正在研究它本身）。

# 2　神奇的大脑

在我们的身体里含有上百亿甚至千亿个神经元细胞和神经胶质细胞，这些细胞构成一个庞大而复杂的信息网络——神经系统。神经系统是对机体内生理功能活动调节起主导作用的系统，分为中枢神经系统和周围神经系统两大部分。而我们最复杂、最神秘也是最引人探索的器官——大脑，正是中枢神经系统中的重要一员（本书之后提到的神经系统，泛指中枢神经系统）。

大脑的复杂性不仅体现在其神经细胞和胶质细胞的数量庞大，更体现在神经纤维间错综复杂的联系。想要走近我们神秘的大脑，第一步最好还是从它的一些基本知识入手，比如大脑是由什么构成的？构成大脑的小家伙们又是怎样通信的？相信在学习了本章有关大脑神经基础及其信息交流的知识之后，你会对大脑有更好的理解。准备好了吗？开往大脑的列车即将出发！

## 走进大脑的微观世界

在学习一门知识或探索一种事物时，我们通常会从它的基本单元入手。同样的，科学家在研究大脑时也通常采用这种“还原主义”策略，由点到面，通过局部进而了解整体。在这一小节中，我们就从微观的角度来简单了解大脑。

如果将大脑比作一座繁忙的城市，那么大脑中最小的工作单元——神经元就是在城市中安居乐业的居民，它们在这座“城市”中井然有序地辛勤工作。除了神经元之外，我们的大脑中还有一种默默无闻的细胞——胶质细胞，它们好比城市中的后勤保障系统，对大脑的正常运作起着重要的作用。有的胶质细胞组成了网络，能够把神经元固定住；有的则担任着巡逻与修护的职能，为神经元的正常工作保驾护航，它们各司其职又相互合作，组成了一座和谐友好的“模范城市”。接下来，就让我们一起钻进大脑，看看这座城市里的“人们”都在忙些

什么吧。

## 神经系统的居民：神经元

其实，将大脑比作一座繁忙的城市并不恰当。别误会，这里并不是指辛勤工作的神经元们不够繁忙，而是大脑中的“居民”数量已经远远超出了“城市”的级别。实际上，整个地球上的人口数都远不及一个大脑中神经元的数量。按照数量级来说，科学界一般认为，人脑中有1000亿个神经元。曾经有人比喻“假如1个神经元是1秒钟，也就是秒针滴答一小格，要想把人脑的神经元都滴答一遍，至少需要3100年”。也就是说，如果1秒数1个神经元，那么我们需要从商朝商纣王开始，不眠不休地数到今天，才能把大脑里的神经元都数完一遍。这可是个天文数字！而这仅仅只是大脑中的神经元数量。另外，神经元们并不仅仅居住在大脑中，它们还生活在神经系统的其他部位，比如脊髓、眼睛、耳朵等，跟随神经系统遍布我们的全身。

那么这些居民都长什么样子呢？一般来说，神经元主要由细胞体和突起组成，根据形态不同会对应不同的功能。图2-1所展示的就是神经元细胞的基本结构，其中细胞体主要负责维持神经元的新陈代谢，是神经元的“本体”，由细胞膜、细胞核、细胞质、线粒体、核糖体等结构组成。神经元突起是由神经元细胞胞体延伸出来的细长部分，根据形态和功能不同，神经元突起可分为树突和轴突。

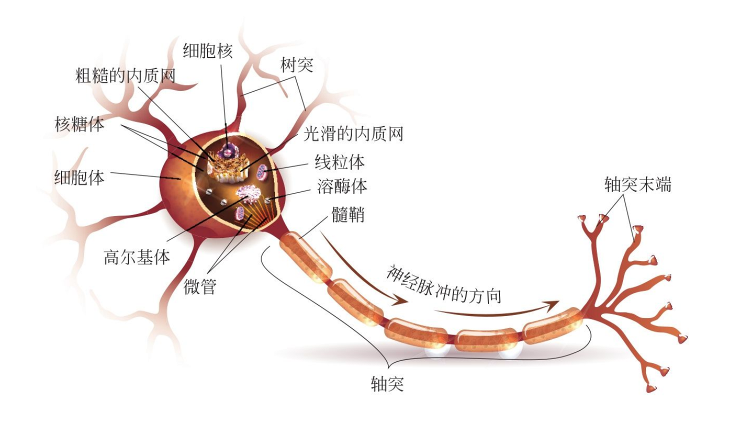


图2-1　神经元细胞的基本结构

树突是从胞体发出呈放射状的一到多个突起，“突”如其名，有的仿佛树梢的枝丫，有的仿佛海底的珊瑚，能够接收其他神经元的轴突传来的信号。轴突长而分枝少，粗细均匀，常起于轴丘，像一条光纤似的将信息传递给其他神经元。神经元细胞之间接收信号的部位称为突触，正如刚才所说，树突接收其他神经元的轴突传来的信号，故而树突又被称为突触后，轴突被称为突触前。大多数神经元既是突触前也是突触后，当它们的轴突与其他神经元建立连接时是突触前，当其他神经元与它的树突建立连接时是突触后。关于突触的详细介绍，我们将会在后文进行。

图2-1所描绘的神经元是以脊髓运动神经元为模型的理想化神经元，实际上神经元具有多种形式。形态学相似的神经元倾向于集中在神经系统的某一特定区域，且具有相似的功能。我们虽难以做到了解神经系统中成百上千万亿的神经元细胞，以及它们各自对脑功能有何独一无二的贡献，但将大脑内的神经元分门别类，探究不同类别神经元的功能却是可行的。

解剖学家根据神经元突起的数目，将神经元分为3种或4种大类。如图2-2所示为3种常见类型的神经元：单极、双极和多极神经元。单极神经元只有一个远离胞体的突起，这个突起能分支形成树突和轴突末梢，常见于无脊椎动物的神经系统。双极神经元，顾名思义，就是具有两个突起的神经元：一根树突和一根轴突。这类神经元主要参与感觉信息加工，例如，眼视网膜的双极神经元。多极神经元具有发自胞体的一个轴突和若干个（至少两个）树突。它们广泛分布于神经系统的多个区域，参与运动和感觉信息加工，是大脑中数量最多的居民。我们通常所说的脑内神经元一般就是指多极神经元。

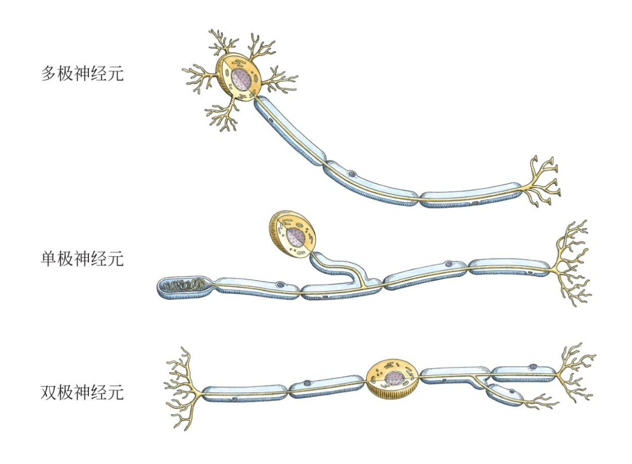



图2-2　不同神经元的细胞结构

除了神经元突起的数目，我们还可以按照树突的形状，神经元之间的联系、轴突长度等对神经元进行分类，感兴趣的话可以自行查阅了解。总之，由于神经元胞体的形状、大小、所处位置以及突起分支的数目、长度、模式、形状、大小等各有不同，科学家们很难对神经元做出统一的分类，只能说“萝卜青菜各有所爱”，不过，其中接受度较广的还是我们刚才介绍的两种以突起的特征来划分的分类方法。

## 神经系统的后勤：神经胶质细胞

我们的生活离不开他人的帮助与服务。类似地，生活在大脑中的神经元细胞能够正常执行功能，背后也离不开后勤细胞——神经胶质细胞的默默服务与保障。神经胶质细胞的数量远多于神经元细胞，大约是其10倍，占脑容量的一半以上，相当于每一个神经元细胞背后配备10个胶质细胞来“服务”它。胶质细胞的英文名是glial cell, glial来自于希腊语中“胶水”一词，原因在于19世纪的解剖学家相信神经系统内的胶质细胞主要发挥着结构支持的作用。的确，有些胶质细胞像胶水一样，组成一个结构网络，能够把神经元固定住；但是也有一些胶质细胞不那么“安分”，在大脑中来回游走，担任着“城市”的监视与修护工作。

如图2-3所示，中枢神经系统内主要有3种胶质细胞：星形胶质细胞、少突胶质细胞和小胶质细胞。星形胶质细胞呈圆形或放射对称形状，是胶质细胞中最大的一种。它们围绕着神经元并与脑血管紧密连接，在中枢神经系统与血液之间构建了一道血脑屏障。这道屏障可以选择性地控制血浆中的溶质通过，阻挡某些血液传播的病原体或过度影响神经活性的化学物质进入神经系统，保持大脑的环境稳定。它仿佛一条把关严格的护城河，在保护中枢神经系统中发挥着至关重要的作用。然而，血脑屏障的这种不完全通透性也给医学界带来了一些考验。如帕金森病，它是一种由于大脑中产生并运送多巴胺的神经元变性死亡，导致大脑多巴胺缺失的严重运动障碍疾病。患者的症状常为静止时手部抖动、肢体僵硬、走路时不能及时调整姿势等。由于多巴胺不能通过血脑屏障，所以患者并不能直接通过血液注射多巴胺来补充脑内衰竭的多巴胺。不过，聪明的科学家已经找到了解决这个问题的方法：利用血液中的多巴胺前体物质——左旋多巴进行治疗。左旋多巴能够穿过血脑屏障，从而被神经元摄取可以转化为脑组织的多巴胺。目前，左旋多巴制剂已成为了治疗帕金森病最主要、最有效的手段。

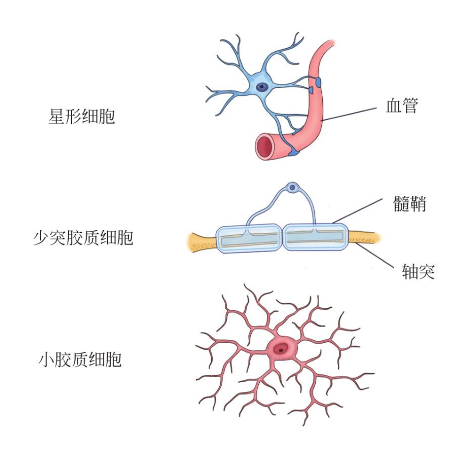



图2-3　中枢神经系统内的神经胶质细胞

少突胶质细胞比星形胶质细胞小，胞突短而少，是中枢神经系统中形成髓鞘的“主力军”。髓鞘是包绕在许多神经元轴突外的脂类物

质，在神经元的生长发育过程中，以同心方式缠绕轴突从而形成髓鞘。它们仿佛电线外的绝缘层，将中间的轴突保护起来，从而保证轴突内电流的传递。此外，有髓轴突的髓鞘被结节分隔成若干节段，由于它们是被一位法国组织学和解剖学家路易斯-安东尼·郎飞（Louis-Antoine Ranvier）发现的，因此被称为郎飞结（nodes of ranvier）。郎飞结对神经元的信号传递具有重要意义，关于这一部分，我们将在后面的章节中进行介绍。

小胶质细胞是一种形状小而不规则的神经胶质细胞，这就是我们之前提到的“不那么安分”的胶质细胞。正常情况下，小胶质细胞处于休息状态，大约以每小时一次的频率与神经元突触发生直接接触，监测突触的功能状态和神经元的活性。当脑内发生炎症、感染、创伤或其他神经系统疾病时，小胶质细胞就会迅速被激活。此时小胶质细胞开始“变形”——胞体增大、突起变短、细胞形态呈圆形或杆状，方便它们“穿墙”赶往受损部位。大量的小胶质细胞抵达目标区域后将进一步调整自身形态呈阿米巴状，以便发挥巨噬细胞作用，吞噬和清除受损的脑细胞。因此，作为城市中的监管者与清道夫，小胶质细胞的形态学改变反映着自身的活化状态，而其活化状态又与脑内受损部位的严重程度密切相关。

神经细胞的交流：神经信号

在前文我们介绍过，大脑这座城市中居住了非常多神经元，但是每个神经元并不是孤零零地独自工作，它们像人类一样需要不停地进行信息沟通和共同合作。那么，神经元和神经元之间是怎么沟通的？打电话？发信息？都不是。人类可以通过无线电波进行沟通，但神经元之间的沟通可不行，它们需要靠真实的物理通路来传递信息。神经元间的交流，有点类似大家小时候玩的“你比划我来猜”的游戏，需要一个一个地将信息传递出去，但并非那么低效，也不是只能一对一地传递信息。

之前我们介绍过，每一个神经元的细胞体周围都有四通八达的突起，如果将神经元的细胞体比作一部座机，那么这些突起就是电话线。轴突负责拨打电话，帮神经元将信息送到外部，树突则负责接通

电话，也就是从其他神经元接收信息。向外拨打电话的轴突只有一条，但负责接电话的树突可以有很多条。依靠轴突和树突，每一个神经元都和数以千计的其他神经元有所连接，联系非常紧密。

当这些“电话线”传递信息时，传导速度可以达到100米每秒，也就是360千米每小时，这和我们高铁的速度接近。所以，其实在我们的大脑中有着不少小小的高铁线，它们正以360千米每小时的速度，有序而高效地传递着神经信号。除了这些飞速的“高铁线”，大脑中有没有速度慢一些的传输线呢？当然有，越细的传输线，传递神经信号的速度就越慢。最慢的传输线，信息传递速度大约是0.5米每秒，相当于一小时只能绕400米标准操场走四圈半。

不过，以上这些说法只是类比型的介绍，要想真正了解神经细胞的信息交流，我们还要从信号的产生说起。简单来说，神经元能够接收外界刺激，这些刺激可以是物理形式的，如眼睛接收的光线、耳朵听到的声音、皮肤感觉到的触摸、电突触收到的电信号等；也可以是化学形式的，如神经递质，或环境中能够使人产生感觉的气味分子等。这些刺激会引起神经元细胞膜的变化，导致神经元膜内外的离子发生流动，从而产生动作电位。在大多数情况下，动作电位的结果是产生一个沿轴突下行传播到轴突末梢的信号，在那里，最终引起突触神经递质的释放。

我们都知道神经元细胞膜是一个磷脂双分子层的结构，如图2-4所示，托两侧圆圆的脂质分子的福，细胞膜能够在细胞内外的水环境中保持完整并控制水溶性物质的进出。就像血脑屏障一样，细胞膜也对经过它的物质有着选择性透过的权力。对于离子、蛋白质和其他溶于细胞内外液体的物质而言，细胞膜它就是一道屏障，毕竟能溶于水的物质都不能很好地溶于细脂质，因此它们便不能轻松地进出神经元细胞体。

动作电位的产生

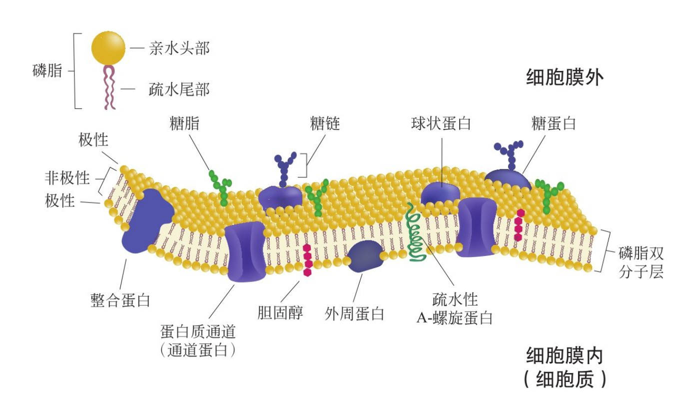



图2-4　细胞膜结构图

神经元细胞膜有两种状态，一种是无事发生，保持静息电位的状态；另一种则是受到刺激，产生动作电位的状态。而神经元的日常，就是在这两种状态间“反复横跳”，以完成受到刺激后在其内部传递信号的工作。在无事发生时，静息状态下的神经元细胞膜两侧存在着一个静息膜电位（-70毫伏，即膜内比膜外电位低70毫伏），为动作电位的产生“随时待命”。当神经元受到刺激时，细胞膜在原有静息电位基础上产生一次迅速且短暂的向周围和远处扩散的电位波动，这种电位波动被称为动作电位。

动作电位的传导

前文提到静息态的膜电位就是外正内负［图2.5（A）］。实际上，某区域产生的动作电位会影响到周围正处于静息态的区域。这很好理解，因为兴奋区域的细胞膜内外两侧的电位差会发生暂时的翻转，即由外正内负转为内正外负，与周围的静息膜之间形成电位差，从而产生局部电流。如图2.5（B）中所示，在膜内侧，电流从静息膜流向兴奋膜；在膜外侧，电流由兴奋膜流向静息膜，结果使原先静息区域的细胞膜内外发生同样的电位变化。因此，在图2.5（C）中可以看出，所谓动作电位的传导其实就是兴奋膜向后移动的过程。

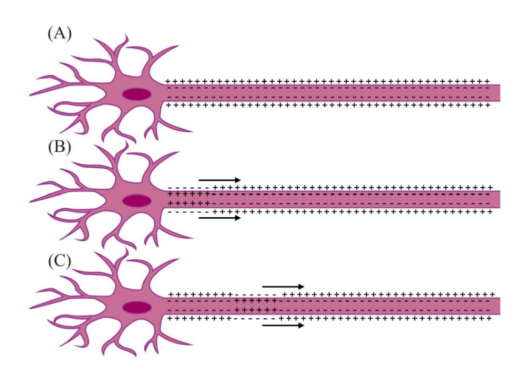



图2-5　动作电位传导示意图

你可以将这种传导过程想象为击鼓传花：花的传递代表了兴奋沿轴突的下行传递。然而，如果传花的人彼此间距离很远，则有可能出现不能把花及时传递下去的情况。类似地，如果下一个具有电压门控离子通道的轴突部分距离太远的话，电流长途跋涉到达该处时，可能已经衰减得无法启动动作电位。那么，这个问题要如何解决呢？答案很简单，只需要尽可能地减少电流在长途旅行中的衰减就可以了。

试想一下，对于一条水管，其粗细会影响水流的流动，管壁的密闭程度也会影响水流在管内流动的距离。同样，神经元细胞的轴突也可以看作一条水管，动作电位的传导就好像水的流动，轴突内的电阻相当于水流在水管中流动受到的阻力，细胞膜的电阻大小对应于水管管壁的密闭程度。可见，通过降低轴突内电阻或增加细胞膜电阻，可以使电流流动更为有效，也流得更远。那么，具体该如何做？

正如增加水管的直径那样，降低轴突内部电阻最有效的方法也是增加其直径，较大直径的轴突可以使传递动作电位更为迅速。但是，对于大型动物，尤其是长颈鹿来说，为了逃避捕食者，要实现足够快地从大脑至后肢运动神经元的信息传递需要多大的轴突直径？答案是非常大。此外，肌肉的控制需要许多运动神经元的驱动，而所有的运动神经元还要包裹于脊髓内，再加上数量多于神经元的神经胶质细胞，长颈鹿的脊髓将会变得非常粗，这显然是不科学的。因此，大型动物想要生存，就必须用其他方式解决这个难题。

髓鞘解决了这个难题，而这也是轴突在保证动作电位有效传导时

采取的另一种方法——包绕在神经元轴突周围的髓鞘提高了细胞膜电阻。髓鞘是以同心缠绕的方式包绕在轴突周围的多层脂质结构，它仿佛电线外的绝缘层，使电流能够沿轴突传递得更远。电流沿着有髓鞘轴突向下快速传递，最终只在髓鞘中段的郎飞结处产生动作电位。因此，看起来动作电位似乎是从一个郎飞结跳到另一个郎飞结，这种传导被称为跳跃式传导。通过这种传导方式，哺乳动物的神经能以120米每秒的速度传导信号，相当于3秒左右就能绕400米操场一圈，这是相当快的速度！

突触传递

前面我们介绍的只是信号在神经元内部的传递，而神经元之间的交流才是大脑有条不紊工作的基础。要实现这一点，神经元之间必须要传递信号。之前我们将神经元间的交流比作通过电话线打电话，实际上，神经元间的信号传递是通过突触完成的，这种传递方式被称为突触传递。突触有两种类型：化学突触和电突触。

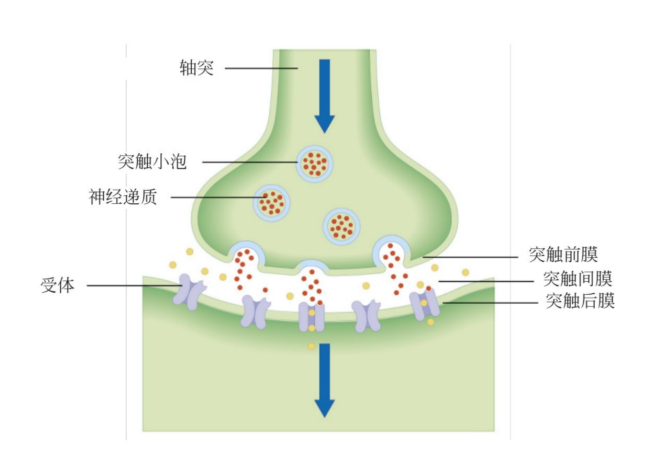



图2-6　化学突触基本结构

化学突触是最常见的突触种类，如图2-6所示，化学突触由突触前膜、突触后膜和突触间隙组成。在多数情况下，一个动作电位到达突触小体所在的轴突末梢，该动作电位可引起轴突末梢去极化，导致$Ca^{2+}$通道打开，使$Ca^{2+}$内流。$Ca^{2+}$在细胞内发挥信使的作用，即可以通过几步中间的生化步骤使信号放大。在此过程中，突触小体中的突

触小泡包裹着神经递质向突触前膜移动，像船停靠在码头一般着位于细胞前膜的蛋白质上。随后，细胞内$Ca^{2+}$浓度的增高导致突触小泡与突触前膜融合，并将其中包含的神经递质释放至突触间隙。神经递质在突触间隙内扩散至突触后膜，如图2-6所示，与嵌在突触后膜的受体结合后，改变突触后膜的离子通透性，使其电位发生变化。

受体是一种化学门控离子通道，一种受体只会与一种物质（即配体）结合，就像一把钥匙开一把锁那样，当受体的“锁”被打开后，离子通道便开放了。通过与配体的特异性结合，受体使细胞在充满无数生物分子的环境中，辨认和接收某一特定信号。

与突触后膜结合后剩余的神经递质最终都去哪了呢？答案是被清除了。若不这样做，它将持续对突触后膜产生影响，导致突触后膜持续兴奋或抑制，这样会导致神经元要么“被累垮了身子”，要么“开始郁郁寡欢”。清除神经递质的方式主要有3种：第一种是突触前膜末梢的主动重摄取，通过一种跨膜蛋白质作为媒介，将神经递质泵回到突触前膜内；第二种是突触间隙内的酶降解；第三种是通过扩散使其远离该突触或作用区域。具体细节会在高中阶段的课本中学到，这里不再展开。

电突触传递信号与化学突触最大的区别在于，电突触没有突触间隙，两个神经元之间由细胞膜相互接触。电突触传递时依靠电流通过细胞膜引起膜电位变化，突触前神经元的动作电位到达轴突末梢，产生局部电流，引起突触后膜膜电位变化，从而引起突触后神经元的动作电位。

通过本小节的学习，相信你已经对大脑内部的“悄悄话”——神经元信号有了一定了解。没错，大脑在获得、传递，包括后续的处理信息时，使用的其实就是电信号和突触间的化学信号。

## 大脑的宏观组成

大脑这座神奇的“城市”作为人体最重要的中枢，无时无刻不在进行着神经元的信息交流。2000年，诺贝尔生理学或医学奖得主、哥伦

比亚大学教授埃里克·埃德尔（Eric Kandel）在他的《神经科学原理》一书中说：“我们整个人类行为的复杂性，与其说是建立在不同的神经元上，不如说是依赖于神经元与神经元之间组成的这种具有精确功能的神经环路和神经网络上。”事实上，我们大脑中有千亿个神经元，每个神经元都和超过1000个其他神经元有所连接。试想一下，由它们构成的神经连接以及各种尺度的神经网络，将是多么庞大的数量级！接下来，我们便从宏观角度切入，认识人脑的整体构造，从另一个角度理解这座“城市”。

## 看一个大脑，分几步?

现在，作为参观者的我们已经钻出了大脑，恢复了日常观察事物的尺度大小。那么接下来，我们就从宏观角度观察大脑的组织结构。

为了看见一个大脑，需要分几步？我们需要穿过哪些人体组织？首先，脑袋的最外面是头发，在头发的覆盖下，是我们的头部表皮层和皮下组织，也就是我们平时说的“头皮”。在头皮的包裹下，是保护着我们大脑的一个很结实的容器——颅骨。颅骨之下是3层很重要的膜，从里到外分别是：硬脑膜、蛛网膜和软脑膜，如图2-7所示，可以看到，我们的大脑在这几层膜的包裹下装在颅骨里面，就像在一个碗里装了一块被保鲜膜包裹着的豆腐。而“碗壁”与“豆腐”间还有很多空隙，这些空隙中填充了一种液体——脑脊液。脑脊液的第一个作用，是为我们的大脑提供缓震，不要小看这个作用，如果没了脑脊液，可能我们一些最日常的头部活动，如一个猛转头，大脑就会因为惯性而在颅骨里面撞来撞去。脑脊液的第二个作用，则是为脑神经组织提供营养代谢等功能。当我们一层一层地穿过这些组织，最终呈现在眼前的便是一个有些粉嫩的、正在轻微搏动的大脑，它提醒着我们这是一个活生生、有思维、有情感、有自我意识的神奇器官。

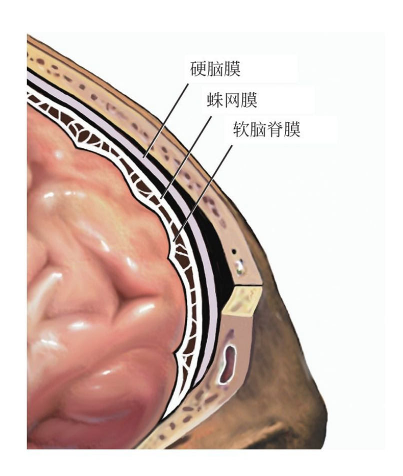



图2-7　脑的被膜

如图2-8所示，大脑由前脑、小脑和脑干构成。前脑包括端脑和间脑，由大脑左右半球和间脑组成，是人脑最大的结构。端脑包括两个左右对称的大脑半球和胼胝体，也是我们看到“大脑”时最吸睛的部分。细心的你也许会注意到，在大脑左右半球的表面，有许多凹凸不平的褶皱，这些褶皱的表层我们通常称为大脑皮质，又称灰质。大脑灰质主要由神经元胞体构成，可以说，大脑大部分神经元的胞体都集中在灰质上。虽然西瓜青色的表皮是我们不怎么吃的地方，好像也没什么用处，但是对大脑来说，大脑皮质却是大脑活动最为密集的地方，是人类思维活动的物质基础，也是调节机体所有机能的最高中枢。在灰质深处有着与之相对应的大脑白质，它主要是由神经元中负责“打电话”的轴突构成，其主要功能就是把一个灰质皮层区域的信息，传输到另一个灰质皮层区域中去，类似于连接不同城市的高铁线路，可以把不同城市的信息进行有效的传输。

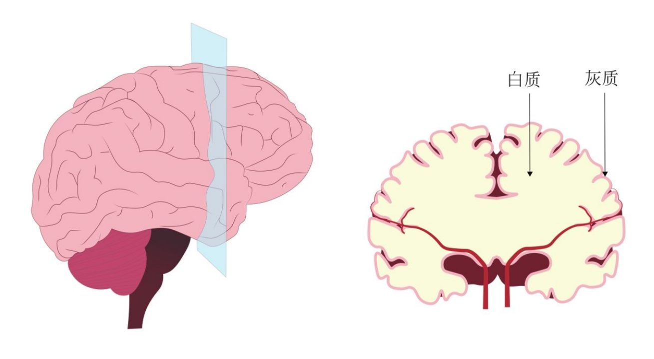



图2-8　大脑灰质与白质

间脑主要由丘脑和下丘脑组成。如图2-9所示，丘脑位于间脑背侧，是皮质下区域和大脑皮质之间传递信息的主要结构，向大脑皮质传递各类感觉信号，具有调节意识水平、睡眠和警觉性等多种功能，被认为是皮质下核团与大脑皮质信息传输的通道。下丘脑位于大脑基底部，对自主神经系统和内分泌系统十分重要。它可以合成和分泌某些神经激素，这些激素反过来刺激或抑制垂体激素的分泌。此外，下丘脑还负责自主神经系统的其他功能，参与调动机体应激反应。例如，当人受惊吓时，下丘脑协调躯体本能的应激反应，调节自主神经系统，促使心率上升、骨骼肌供血增加。而在休息时，下丘脑调节自主神经系统，在确保脑的营养基础上增加肠胃蠕动，促使血液进入消化系统。此外，下丘脑还具有控制体温、饥饱、依恋行为、口渴、睡眠和调节昼夜节律等功能。

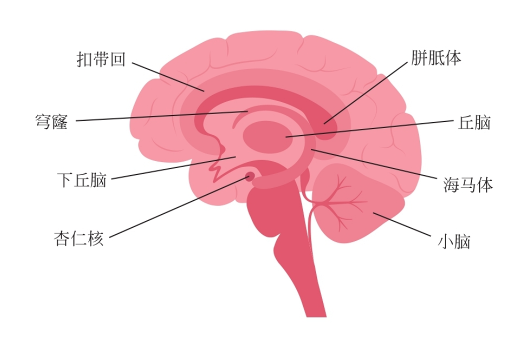



图2-9　间脑解剖示意图

小脑，顾名思义为小的大脑，如图2-9所示，是一块覆盖于脑干结构上部、处于脑桥水平位置的神经结构。小脑虽然只占大脑总重量的10%，但其包含的神经元和回路比大脑的其他部位都要多。小脑能够调节运动控制，维持身体平衡，它并不直接控制运动，而是通过整合有关身体和运动指令的信息来调节运动。除此之外，还参与一些认知功能，如注意力和语言等。

脑干位于大脑后部，自下而上由延髓、脑桥和中脑组成。延髓部分下连脊髓，与脑桥共同构成后脑。从脊髓到延髓、脑桥、中脑、间脑、大脑皮质，大脑结构和功能变得越来越复杂，但这并不意味着脑干的功能仅仅是辅助性的。脑干处于脊髓以上中枢结构的最底部，起着“承上启下”的作用，所有从身体传递到大脑和小脑的信息都必须经过脑干。除此之外，脑干还具有控制心血管系统、呼吸、疼痛、警觉性、意识等与生命相关的功能。因此，脑干损伤是一个非常严重，甚至会危及生命的问题。

## 大脑地图

众所周知，地图是按照一定的法则，有选择地以二维或多维形式在平面或球面上表示某地区若干现象的图形或图像，能够科学地反映出自然和社会经济现象的分布特征及其相互关系。假设我们周末想要去一家新开的商场消费，首先就需要通过地图准确地得知它的定位。而我们的大脑，也是有“地图”的。有人将大脑这座城市的地图分为了3种模式，分别是行政地图、道路地图和功能地图。接下来我们就来了解大脑这份地图要如何使用。

行政地图：脑叶

行政地图就是区划图，如北京市有海淀区、朝阳区、昌平区等，大脑也有不同的分区，我们将其称为脑叶。如图2-10所示，根据大脑的解剖学分区，可以将其主要分为4个叶：额叶、顶叶、颞叶和枕叶。如果将边缘系统称为边缘叶，就是5个叶。这些脑区的名字来自于对应颅骨的解剖位置。额叶是额头部分对应的脑区；顶叶是头顶的区域；颞叶对应我们耳朵上方的区域；而枕叶则是我们平躺睡觉时，

后脑勺枕枕头的区域。

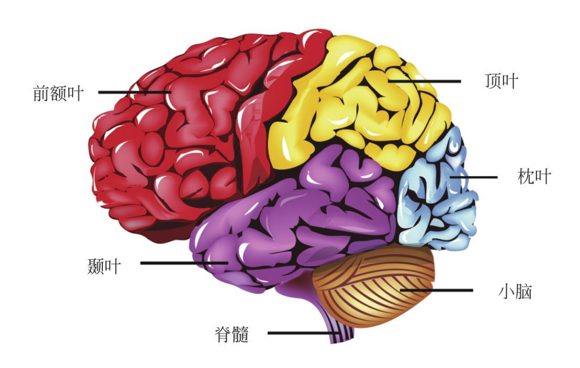



图2-10　大脑的脑叶划分

划分城区时，不同的区域承载着不同的城市功能，如北京海淀区是北京的科技和教育聚集地，上海浦东新区是上海的金融贸易中心。类似地，大脑中的脑叶划分也是如此。额叶与躯体运动、语言和各类认知加工以及情绪管理功能有关，擅长进行信息的深度加工和预测。顶叶中的初级躯体感觉和次级躯体感觉区接收痛觉、触觉、温度感觉以及本体感觉等信息，并在顶叶中把不同的感觉信息整合在一起。枕叶是最小的脑叶，擅长处理视觉信息。颞叶包括听觉、视觉和多通道加工区域，还有与语言有关的皮质，该区域损伤将导致失语症。此外，颞叶还与知觉和记忆功能相关。边缘叶位于大脑内侧颞叶下方，属于大脑的“城郊区域”，由扣带回、下丘脑、丘脑前核、海马、杏仁核、眶额皮质和部分基底神经节构成，参与人脑的情绪、学习和记忆的加工。

你也许会对海马和杏仁体有所耳闻，尽管它们听上去似乎与大脑有些格格不入，实际上它们的名字来源于其形状，如图2-11所示，是一种“象形名字”。海马，是一个形似海马的结构。它能够短暂存储外界信息，将其传输至皮层，形成对刺激的长期记忆。该区域参与学习、记忆过程，并在空间记忆和情境记忆中起着重要作用。杏仁核是形如杏仁状的核团，参与情感反应和社会性过程。如果你每次在体测或者考试前都会感到焦虑不安、手足无措，这并不是因为你将这种恐惧“刻进了DNA里”，也许是因为你对这些事情的负面情绪已经“刻

进”了杏仁核里。



图2-11　海马（红）解剖示意图

道路地图：沟回

图2-10中我们可以根据颜色清楚地区分各个脑区，但是对科学家来说，他们是如何对脑叶进行划分的呢？在我们的城市里，各个区的划分，往往是以某条路作为界限，比如路的这边是海淀区，路的那边是朝阳区。脑叶的划分也类似这样，是以一些显著的结构来进行划分的。前文中我们讲到，大脑的表面有凹凸不平的褶皱。人们把凹进去的地方叫作脑沟，凸起来的地方叫作脑回，这些错综复杂的沟回就像城市地图上的道路一样，往往可以作为我们定位脑区的参照。

那么，这些沟回都是如何被命名的呢？平常我们在命名道路的时候，经常会使用区域加方位，比如中关村南大街、学院北路、北三环西路等，大脑沟回的命名也是这样。不过在大脑中显然没有东西南北这种说法，我们常用“上下”“前后”“内外”及“腹侧和背侧”来描述大脑中的相对方位，如图2-12所示。“上下”很好理解，靠近头顶的就是上，靠近脖子的就是下；“前后”也很好理解，靠近额头方向的是前，靠近后脑勺方向的是后；而“内外”是以大脑的中心轴为参照，靠近鼻子的就是内侧，靠近耳朵的是外侧；至于“腹侧”则是靠近我们肚子所在的一侧，“背侧”是靠近后背的一侧。现在，掌握了“上下”“前后”“内外”及“腹侧和背侧”，我们基本上在大脑里就不会迷路了。可以通过几个例子来实践一下，例如：额上回是额叶中最靠近头顶的脑回；外侧

枕回是枕叶中靠近耳朵的部分等。同样的，大脑皮质区域也可以利用方位进行描述，如之前提到的前额叶皮质，就是额叶中最靠近额头的区域；而内侧颞叶则是颞叶中最靠近大脑中央的区域等。

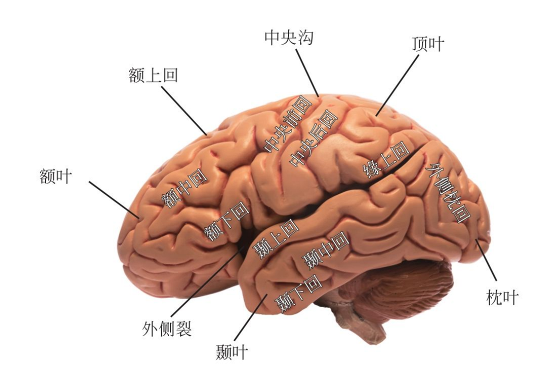
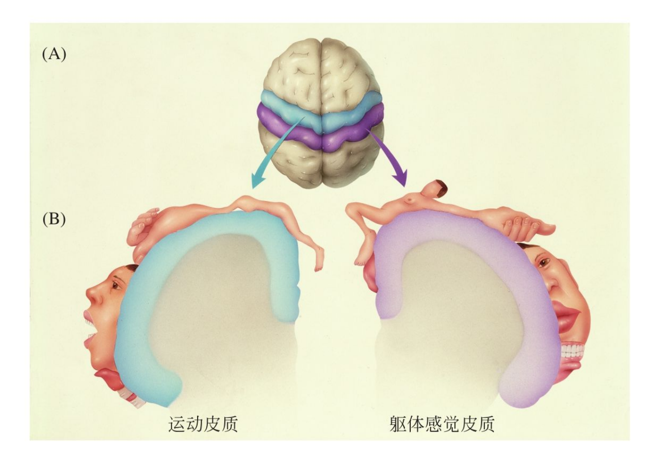



图2-12　人类大脑皮质的左视图

长安街是北京家喻户晓的主干路，同时也是横贯北京城区的东西中轴线，其实在我们大脑纵横交错的道路中也有这么一条“长安街”，它的名字叫作中央沟。如图2-13所示，中央沟几乎从中间纵贯我们的大脑两侧，是额叶和顶叶的分界线。其周围凸起来的部分我们称之为中央前回和中央后回，这两条脑回对我们日常生活非常重要。中央前回是我们的初级运动皮层，所有对身体不同部位运动的控制都是由它参与完成的；中央后回是躯体感觉皮质，顾名思义，我们身体上的大部分感觉都由这部分感觉皮层来处理。

20世纪40年代，怀尔德·彭菲尔德（Wilder Penfield）医生和他的同事利用脑外科手术的机会研究了病人在清醒的情况下直接刺激皮质的反应，并发现了病人的身体表面和上述这两片皮质区域之间有一种地形上的对应关系。通过把身体各部分画在运动和躯体感觉皮质的冠状切面上，他们得到了一张著名的图——侏儒图，如图2-13所示。从图2-13中可以看到，我们主要身体部位的实际大小和该部位皮质表征的大小之间并没有直接的关系。如运动侏儒的手指、嘴部和舌部肌肉远大于正常人的身体比例，这表明当我们在操纵物体和说话时有大片皮质区域参与精细调节。




（A）中央前回（蓝）与中央后回（紫）；（B）大脑皮质侏儒图

图2-13　中央回示意图

功能地图：脑区

我们知道，人脑有许许多多功能，不仅能够处理来自感觉器官的各种信息，还能够进行运动控制、认知控制、产生情绪和情感等高级认知活动。这些不同的心理体验以及行为背后，对应的大脑活动也不尽相同。对于这些大脑活动的描述，光靠大脑的结构图是不够的，我们还需要一张大脑功能地图。

人类尝试去绘制大脑功能地图，已经有100多年时间了，学者们按照多种方式对皮质进行分区，哪怕是现在，这个绘制活动也还在进行中，在这些分区中应用最广的是布罗德曼分区。20世纪初，科比尼安·布洛德曼（Korbinian　Brodmann）通过分析细胞和组织形态之间的差异，将大脑皮质划分为52个代表不同功能的区域。不过，经过研究者们多年的探索与修正，现在的标准版本比布罗德曼的最初版本省略了一些脑区，如图2-14所示，不同区域执行不同的功能。例如，布罗德曼1区（一般称作BA1）表示初级躯体感觉皮层，BA4是初级运动皮层，BA17为初级视觉皮层，BA41和BA42为初级听觉皮层等。

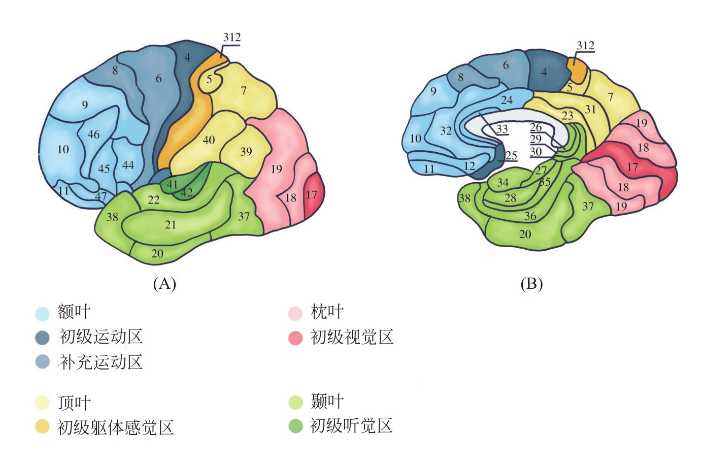



图2-14　左半球侧面观（A）与右半球内面观（B）的大脑Brodmann分区

尽管不同的科学家可能对大脑的功能区域划分不同，但是有几条规律却是大家都心照不宣地认可的。

第一，大脑中的“居民们”非常团结互助。即使是一个最基本、最简单的大脑功能，也不是由单独一个神经元完成的，往往是多个神经元一起来做同一件事情。

第二，大脑中的“居民们”也存在聚居地。承载着相同或类似功能的神经元，以及神经元群，往往扎堆生活在一起，也正是基于这条规律，才有了“脑功能区”的概念。

第三，大脑也有“互联网”，通过这个网络，居民们共住“大脑村”。近几年，越来越多的研究者发现，大脑比较复杂的功能，往往是由几个小的功能区之间协作完成的。每个脑功能区不仅要完成自己的工作，还要不停地和其他脑功能区交换信息、沟通工作进展、分享工作成果。这些不同脑功能区之间的交流，就形成了脑功能网络，而这个脑网络中每一个脑功能区，就是这个网络中的一个节点。

总而言之，人类的心理活动和行为千变万化，如何绘制更精细的脑功能区图谱，为人类探索大脑提供更好更准的地图，是现在脑科学领域不断追求的目标之一。

## 大脑的可塑性

你认为人类大脑会发生变化吗？

2005年《科学》杂志上的一篇文章报道，科学家发现一个和人类大脑大小发育相关的基因，这个基因被称为“异常纺锤形小脑畸形症相关基因”，它最近的一次更新，是在5800多年前。正是由于这个基因的更新，人类才产生了较大的脑容量，并获得了较好的认知和学习能力。5800多年前是什么概念？要知道，我们常说中华文明上下5000年，和它的更新时间相比竟然还少了将近1000年的时间。换句话说，现代人类出生时自带的大脑，它的“预装系统”和容量大小其实与炎帝、黄帝、孔子、孟子、秦皇汉武等人并没什么太大的区别，我们都在使用最新版本，是不是感到很惊讶？

但是你可能会想：我虽然不一定会比孔子有学问，但我至少会比和他同一时代的普通百姓聪明吧？请放心，虽然大家的“预装系统”都是一样的，但是在这个基础上，由于我们接触到的外部环境、社会环境、受教育的程度、人生经历的不同等这些复杂的因素影响，我们的大脑在不断发生着或大或小的改变。也就是说，接受过现代教育的你，一定是比古代的普通百姓聪明得多。

在“大脑”这一部分中，我们已经学习了许多关于大脑结构和功能的知识，这些知识都是静态且统一的。但是每个人都有独一无二的大脑，哪怕是同一个人，大脑在他青少年时期也会与老年时期有所不同。“大脑在人们一生中是在不断动态变化的”这件事，基本上是目前科学家比较公认的看法了。

例如，我们并不是天生只会一种语言，人类大脑的“预装系统”让我们具备了习得任何一种人类语言的能力，只不过安装进去的是哪种语言，就和你的成长环境及语言学习经历有关了。不同的语言学习经历，确实也会对大脑有着不同的塑造。科学家们对不同母语人群的大脑进行了研究，对比了以汉语为母语的人和以英语为母语的人他们大脑皮层中和声调加工相关脑区的差异。我们都知道，在汉语中，同一个字用不同的音调说出来，字义就完全不同。例如，妈、麻、马、骂，都是ma这个音，但是声调不同，意思也不相同。但在英语里面，如apple？apple！无论音调怎么变，apple还是苹果的意思。由此可见，说汉语的人，在日常语言交流中，对声调的加工显然要比说英语

的人强烈得多，而这一点在大脑皮层中有很明显的反映。研究发现在和语言声调加工有关的脑区，也就是我们右侧颞叶前部的区域，汉语母语者大脑的灰质集中度，要显著高于英语母语者。你可能会问，这种差别会不会是人种差别导致的呢？研究者显然也考虑到了这种可能，他们还找来另外一组人进行验证，他们都是西方人，且是英语母语者，但是这组人都学过汉语，学龄从3年到7年不等。科学家比较了这组人的大脑，结果发现，他们的右侧颞叶前部和声调加工相关的脑区里，灰质集中度也明显增强了。此外，还有研究者发现，对于双语使用者，他们大脑里的活跃区域面积要比那些只会说一种语言的人大得多，也活跃得多。

上面的例子说明，大脑是可以改变的，也就是大脑具有“可塑性”。人们常常认为大脑的发育存在一个关键期，青春期后大脑就失去了原有的活力。然而近年来，越来越多的研究表明，成人的大脑同样具有可塑性。其实我们的大脑从未停止过改变和调整以达到最优的神经回路，只是在幼年时期这个过程发生比较迅猛。正是因为大脑具有这种可塑性，我们虽然一出生都装配了同一个“预装系统”，但是在成长中，由于环境和个人经历的不同，大脑也在进行相应地变化调整、不断地被塑造，最终形成了每个人独一无二的自我。

## 大脑是如何实现可塑性的？

一般来说，对大脑可塑性的讨论分为两种：一种是毕生发展期间正常大脑根据经验与学习而重组神经路径的终身能力，另一种是脑损伤之后作为补偿功能机制的神经可塑性。

关于第一种大脑可塑性的例子，比如，常年训练小提琴的人比未经过任何音乐训练的人能激活更多对应的躯体感觉皮层，而且激活区域的大小与学习小提琴的起始年龄呈正相关。

神经科学家唐纳德·赫布（Donald Hebb）提出了大脑可塑性的神经机制理论，即突触的可塑性。简单来说，就是突触前神经元向突触后神经元的持续重复地刺激，可以导致突触传递效能的增加。不过，更进一步的研究将刺激分为了高频刺激和低频刺激，对赫布理论进行

了一些补充和修正。

现在我们一般认为突触可塑性主要包括两种模式：①长时程增强，即在短时间内快速重复高频刺激，突触传递效率呈现持久的增强现象，因此称为长时程增强；②长时程抑制：与长时程增强相反，长时程抑制指长时间内重复低频刺激，突触传递效率呈现持久的降低现象。长时程增强可以强化记忆的形成，而长时程抑制则对记忆内容进行选择、确认、核实，二者相互影响，调节大脑的学习和记忆功能。

除了神经元间交流效能的改变能够影响大脑的可塑性，大脑中神经元和神经胶质细胞的表型变化也能够影响其可塑性。例如，经验与学习可以驱动树突棘和轴突的生长或神经突触发生，这可能是神经回路自适应重构的基础。脑损伤可能激发突触可塑性机制，树突棘的数量、大小和形状在损伤后可以发生迅速变化，以促进功能恢复。在神经元迁移、成熟和退化过程中，星形胶质细胞的形态发生了明显变化，表现出高度的表型可塑性，不断适应大脑环境的变化。

上述这些大脑微观层面的变化能够引起宏观尺度上的功能重组，也就是我们说的大脑功能可塑性。在同一区域内同一功能可能由多个皮层区域共同执行，当主要功能部位受损时，可以通过相邻皮质的参与得到补偿，这些皮质与受损区域位于同一区域，损伤后通过代偿机制产生局部超兴奋性。若该受损区域内的再分布不足以恢复功能，则由同一功能网络的其他区域进行代偿。

2021年《柳叶刀—神经病学》发表了一个罕见病例的研究，该病例研究的对象名叫丹尼尔·卡尔，他生下来就患有中风，婴儿时期的严重中风导致他的大脑两侧严重受损。从图2-15中可以看到，他的大脑与其他年轻人的相比几乎缺少了1/4，在与运动、思考、情感、记忆等高级功能有关的大脑区域中，大量组织明显丧失。然而，丹尼尔似乎并没有出现认知、记忆或情感方面的问题，仅在运动技能评估结果中，相对于左上肢，他右上肢的力量、速度和敏捷性都更弱。如今，他已经20多岁，过着非常正常的生活。

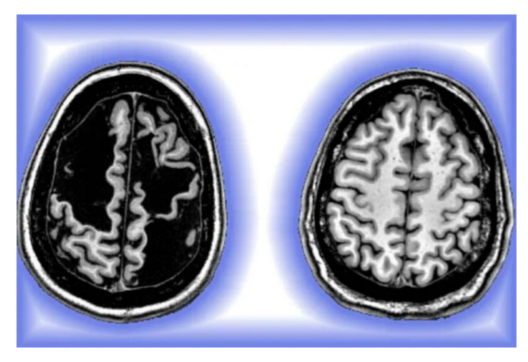


图2-15　丹尼尔的大脑成像数据（左）与正常年轻人的大脑成像数据（右）

## 大脑可塑性的应用

研究大脑可塑性的原理，是为了更好地理解大脑的神经机制，进而为人类攻克神经类疾病提供可能。目前，以改善大脑可塑性为目的的治疗主要包括药物治疗、行为训练、物理调控等手段，下面我们简要介绍这几种方法。

药物治疗

改善大脑可塑性是中枢神经系统药物的重要机制。与大脑可塑性原理对应，药物治疗也可以通过两种途径改善患者脑功能，下面我们用两个例子来简单说明一下。

乙酰胆碱酯酶抑制剂主要用于治疗痴呆，可以缓解与记忆问题相关的胆碱能阻滞，促进海马的长期记忆巩固，诱导海马长时程增强。

目前研究表明，抑郁症患者前额叶皮层中的树突数量和大小都会减少。赛洛西宾是一种新型突破性治疗抑郁症的药物，它能够帮助大脑建立新的树突连接，且这些连接具有较为理想的强度和稳定性。在给药后立即形成的树突连接中，大约1/2的连接在1周后仍然完好无损，大约1/3的连接在34天后仍然完好无损。

行为训练

临床上常使用行为训练改善大脑功能。康复治疗（即行为训练）旨在改善患者的功能和生活质量。利用活动依赖的神经可塑性进行特

定功能的康复可以使康复效果最大化。这一原理可以应用于不同的功能，如运动控制、语言和认知。

使用康复训练对存在阅读及言语障碍的儿童进行治疗便是一个很好的例子。当我们阅读时，大脑需要一直不停改变眼睛运动的指令。当读到句子的一部分后，大脑就会命令眼睛移到句子的后半部分去。但是，患有阅读障碍的儿童无法完成这个动作，他们在阅读时非常慢，同时还会存在漏字、跳行等问题。在康复训练时，通过对阅读障碍的儿童进行行为训练，如用手描绘复杂的线条等，刺激患者前运动皮质区，从而改进儿童在说话、写作和阅读3个方面的表现。同样地，患有言语障碍的儿童在训练早期，快速变化的语音通过放大和重复播放来消除言语歧义。结果表明，经过训练后，孩子们的自然语言理解能力有了很大提高。

此外，有研究发现，利用视频游戏的计算机化程序可以改善视觉感知缺陷，以及与年龄相关的认知功能退化。而如何将一个认知领域的特定任务训练推广到更广泛的功能领域也是行为训练的一大研究热点。

物理调控

物理调控技术是通过神经可塑性的调节来改善大脑功能，主要包括有创和无创两大类。深部脑刺激，俗称“脑起搏器”，是一种经典的有创神经调控技术。如图2-16所示，该技术通过植入电极将电脉冲发送到大脑中特定区域，调节该区域的功能活动，从而改善患者的临床症状。目前，深部脑刺激常用于治疗特发性震颤和肌张力障碍等运动症状，也可用于治疗强迫症和抑郁症等精神疾病。对于不同疾病，脑刺激的电极植入位置也不相同，其中最常见的植入位置是与帕金森病相关的丘脑底核附近。除此之外，人们还在积极研究将其用于治疗抑郁症和其他精神疾病。

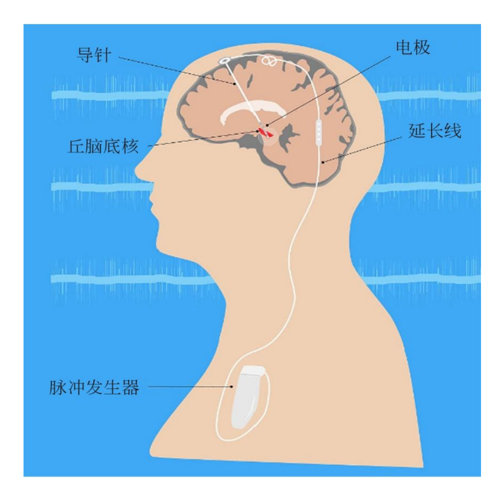



图2-16　深部脑刺激示意图

无创神经调控技术又称非侵入性脑刺激技术，主要包括经颅电刺激、经颅磁刺激等手段。经颅电刺激通过附着在头皮表面的电极施加微弱的电流刺激（一般不超过2毫安）。电流在电极间流动的过程中会穿过头皮、颅骨和脑脊液到达大脑皮质，调节皮质组织区域内神经元膜的极性，进而影响神经元兴奋性、改变神经元活动。经颅磁刺激是一种基于法拉第电磁感应原理，通过外部变化的磁场在大脑中诱导产生电流的无创神经调控技术，因此也被称为“基于电磁感应对大脑进行的无电极电刺激”。在研究大脑机制及疾病治疗中显示出了较好的应用前景。

正所谓“活到老，学到老”，大脑可塑性研究在一定程度上说明了成年后人类仍然可以通过学习和训练锻炼我们的大脑。然而人类大脑可塑性的机制目前还尚未完全明了，学习是如何引起脑活动状态变化的，大脑皮层功能的变化与脑内神经元、突触之间存在怎样的联系，脑发育的关键期和可塑性的关系等一些更深入的问题，目前也都还不太清楚，这些都是将来需要科学家不断探索研究的问题。

此外，对人类大脑的研究和模拟是推动人工神经网络进步的重要手段。关于这部分的内容，将在书中后续章节中进行介绍（详见第5章）。若你对这些问题同样感兴趣，也欢迎你在不久的将来加入进来，共同探索大脑的奇妙。

# 小结

在这一章中，我们首先从微观角度学习了大脑的神经基础。神经元作为大脑中的“居民”，数量可达成百上千亿个，其大部分的活动都在胞体完成，而树突和轴突则是它们用来相互交流的物理通路。神经胶质细胞作为城市中的后勤系统，为神经元的正常工作保驾护航：有的胶质细胞组成了一个结构网络，像胶水一样，能把神经元固定住；有的胶质细胞可以在大脑中移动，担任着“城市”的监视与修护工作。

神经元细胞的信息交流除了依赖于树突和轴突的物理通路，还依赖于神经元间的突触，同时也离不开其动作电位的产生与传导。这些神经元的相互交流和连接组成了不同的神经环路，进而能够使大脑具有宏观的结构和功能。大脑由前脑、小脑和脑干构成，每个部分都承担了属于自己的责任，它们相互协作，缺一不可。

人类大脑具有可塑性，神经元和神经胶质细胞的表型变化能够影响其可塑性，神经元间的交流效率也能够通过刺激而改变。药物治疗、行为训练和物理调控是目前较为常用的3种改善大脑可塑性的手段。大脑可塑性研究在一定程度上证明了“活到老，学到老”背后所蕴含的科学道理，也激励着科学家对其展开更深入的研究与探索。

# 3　脑认知科学的兴起

通过前两章的介绍，我们对脑科学发展的历史及大脑的神经生物机制有了初步的认识。现在，相信你已经对大脑传导机制，如神经细胞在大脑中扮演怎样的角色、信息如何从一个神经元传递到另一个神经元等，理解得比较清楚；对不同脑区的功能划分也有一定的了解。那么，恭喜你！你已经穿越时空，站在巨人的肩膀上看到现代科学家的研究前沿了。

然而，只了解神经细胞如何处理信息是不够的。大脑作为人类最复杂的器官，除了调节人体功能外，也是意识、精神、语言、学习和记忆等高级神经活动的物质基础。这些神奇的高级神经活动即人的脑认知功能，它们不仅决定着人类在大自然中的特殊与高等，也吸引着科学家们对其孜孜不倦地研究。

了解脑认知功能产生的机理，对于人类发展具有重要意义。比如在人工智能领域，想要设计出一个能够根据指令完成动作的机器人，就需要知道人的大脑是怎样处理语言和进行运动控制的。在这一章中，我们将介绍几类经典的脑认知功能——感知觉、运动控制、记忆与注意。

## 大脑的知觉与感觉

你能分清知觉与感觉吗？从科学的角度讲，知觉通常由视觉、听觉、嗅觉、味觉和触觉5种感觉整合得到。在正常人的知觉中，各种感官的相互作用是十分重要的，它们互相协调，构成认知功能的基础，使你能够完整一致地感受这个世界。接下来，我们将简单地介绍各个感觉，以及一些生活中常见现象背后蕴藏的感官系统“秘密”。

### 五感的“主角”——视觉

假如让你失去一种感觉，你最不想失去的是哪种呢？我猜大多数人的选择都会是视觉。人类所感知的外界信息中80%的信息都来自于视觉，并且人类大脑皮层的1/3面积都与视觉相关——视觉，是当之无愧的五感“主角”。

你知道吗，人的眼睛是分主副眼的，在专业术语上称为优势眼和非优势眼，或左/右利眼，就像左/右利手一样。想知道究竟哪只眼睛是自己的优势眼吗？你可以做一个这样的小测试：

首先，将手臂伸直置于胸前，双手外翻并交叠，使两只手的虎口处可以形成一个三角形小孔（越小越好）；其次，通过虎口间的小孔寻找一个远处的物体，可以是墙上灯的开关，或是窗外的路灯，保持两眼同时睁开的状态，移动手臂，使这个物体处于小孔的中央；最后，保持身体不动，轮流闭上一只眼睛。若你看到这个物体仍在小孔中央，那么此时睁开的那只眼睛就是你的优势眼。有研究表明，较多数人都是右利眼，仅有30%～50%的人为左利眼。怎么样，你的优势眼是哪只呢？优势眼的作用主要是单眼视，假如你老花眼了，需要一只眼睛看近，一只眼睛看远，医生就会根据你的优势眼来选择。

测试完优势眼，你可能会想：既然我的大脑更“信任”某一只眼睛，那为什么人还需要两只眼睛呢？实际上，我们的眼睛就像照相机一样，能够感受光线的强弱，但是一只眼睛只能得到一个平面图像，因此只能确定物体的方位，无法判断物体的距离。而当人用两只眼睛注视物体时，双眼分别能获得一个不同位置的物体图像，这两个图像之间存在一定的水平差异，即视差。视差经大脑加工后，便产生了使我们能够感知三维空间各种物体远近前后和高低深浅的立体视觉。

立体视觉并不是天生的能力，当我们还是婴儿的时候，必须通过体验周围的空间去学习这种能力。依靠移动身体、触摸物体以及保持平衡，我们才能在大脑中构建周围环境的立体地图。在消化这些经验的过程中，我们掌握了三维空间特征，从而形成立体视觉。依靠精确的立体视觉，我们才能快速判断物体与自己的距离，能够在复杂环境中安全地移动。

由此可见，视觉是十分重要的。那么我们是如何获取视觉信息，

大脑又是如何接收、加工这些信息的呢？

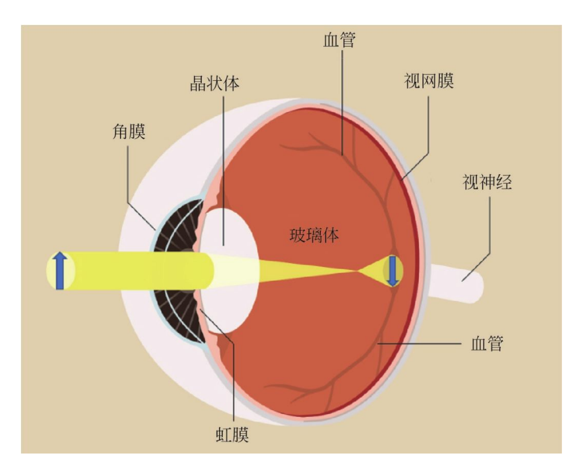



图3-1　眼睛成像示意图

视觉信息包含在物体反射的光线中，由眼睛进行接收和处理。如图3-1所示，光线通过角膜折射和晶状体聚焦后，图像会被反转，然后通过充满眼眶的玻璃体到达眼球的后表面——视网膜。视网膜上的感光细胞将光刺激转换为大脑可以理解的神经信号，从而对视觉信息进行汇聚。最终，视网膜上的另一类细胞——神经节细胞将信号传出，通过视神经传递到中枢神经系统，在那里，大脑处理这些信号并形成视觉。

图3-2展示了视觉信息是如何从眼睛传递到大脑的。可以看到，每个眼睛对应的视神经分为两部分：颞侧（靠近耳朵侧）和鼻侧（靠近鼻侧）。由于光沿直线传播，右视野的光线经过小小的瞳孔，投射在眼球的左侧视网膜上，并由对应的视神经进行传递，即右视野的物体会刺激左眼视网膜的颞侧和右眼视网膜的鼻侧，左视野同理。在这之后，颞侧的视神经分支继续沿着原先的方向前进，而鼻侧的分支则经过视交叉投射到相反的一侧，最终使右视野的所有信息被投射到大脑左侧半球，左视野的所有信息被投射到大脑右侧半球。

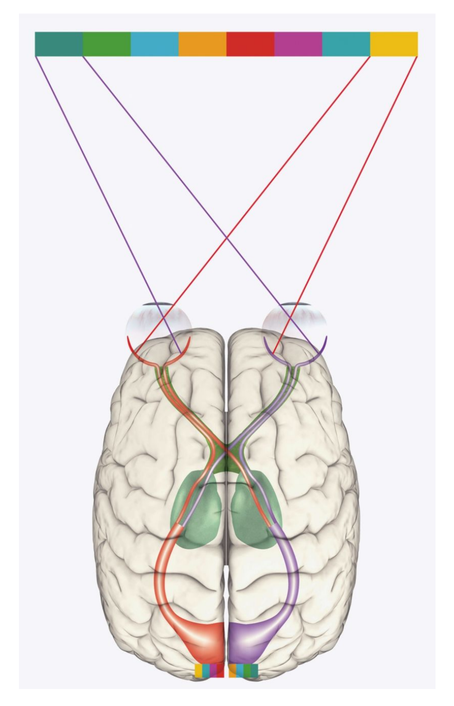


图3-2　视觉系统的初级投射通路

当信息进入大脑之后，根据视神经终止于皮质下结构的位置，视觉通路可以分为视网膜-膝状体通路和视网膜-丘体通路。图3-2所示的就是视网膜-膝状体通路，这条通路包含超过90%的视神经轴突，你可以简单理解为有90%的视神经通过这条路将信息传递至大脑皮层的初级视觉皮层。在这之后，信息再次“兵分两路”：一路由大脑背侧延伸向顶叶，即“向上走”，这条通路负责对运动视觉、空间方位等位置信息的分析，因此也称where通道；另一路由大脑腹侧投射至颞叶，即“向下走”，这条路负责对物体的辨认，因此也称what通道。而剩下10%的视神经则将视觉信息传到了其他皮质下结构，如上丘和枕核。虽然10%听上去并不算多，但由于人类视神经十分丰富，10%的视神经轴突可能就与一只猫的视网膜神经节细胞总数相当。

在这里我还想向你分享一个概念——盲视。你可能觉得这个词非常奇怪：既然已经是“盲”了，怎么还能“视”呢？实际上，“盲视”的意思是眼睛完好，但视觉皮层的某块区域受到损伤，导致患者的大脑虽

然可以接收新的视觉信息，却不能有意识地进行信息获取，换句话说，就是看见了，但意识不到，用一个成语来形容就是“视而不见”。

有趣的是，一些盲视患者也能对盲视视野内的信息做出反应，他们的盲视似乎是相对的。例如，有些患者能说出盲视视野内物体的颜色，但他们不会主动说看见了什么颜色，只有在必须对颜色做出猜测的时候他们才会说出来。造成这一现象的原因可能是盲视患者大多受损的是视网膜-膝状体通路，而并未受损的视网膜-丘体通路仍可以处理一些低级视觉信息。

## 五感的“配角”——听觉

在开始学习这部分内容之前，你可以做个小游戏：找一位朋友，告诉他你将要说一句英文，然后不发出声音，仅用口型说“elephant juice”，说完后让他猜猜你说了什么——他大概率会回答“I love you”。这两句话的口型几乎是一致的，但表达的意思却天差地别，可见听觉作为日常生活中视觉的重要补充，虽然被称为“配角”，但在我们感知世界的过程中也同样扮演着重要的角色，大脑在缺少听觉信息的情况下，很有可能会产生错误的判断。

我们是如何听到远处的声音，并准确地定位声音来源的呢？人类之所以能够识别声音并判断其产生的位置，很大程度上依赖于听觉系统。听觉系统由耳朵和相关脑区组成。耳朵由外耳、中耳和内耳3部分组成，其主要任务是将声音转化为神经信号，为了更好地了解它的工作原理，让我们跟随声音进行一场“耳朵之旅”吧！

声音以波的形式通过空气、液体或固体等媒介传播至外耳，外耳包括耳郭和外耳道，主要起集声作用。如图3-3所示，声波在外耳集合后，由中耳的鼓膜、鼓室和听小骨传递至内耳。鼓膜，顾名思义，是一种与鼓面相似的膜，声波可以使其产生震动，就像敲鼓时鼓面的震动。鼓膜会牵引听小骨一起震动，从而推动耳蜗中的液体流动——这便将声音传到了内耳。内耳由半规管、前庭和耳蜗组成，其中耳蜗是听觉系统的主要部分。耳蜗形似蜗牛壳，是一个内部充满着液体的盘旋骨管。耳蜗内部有一种叫作基底膜的膜状结构，基底膜上覆盖有

特殊结构的毛细胞，它们会随耳蜗中液体和基底膜的震动而摇摆。而这一动作会引发膜电位变化并释放神经递质，使得支配毛细胞的听觉神经产生兴奋和冲动，进而将声音信息传到听觉中枢，最终由大脑皮层解析识别，形成听觉。

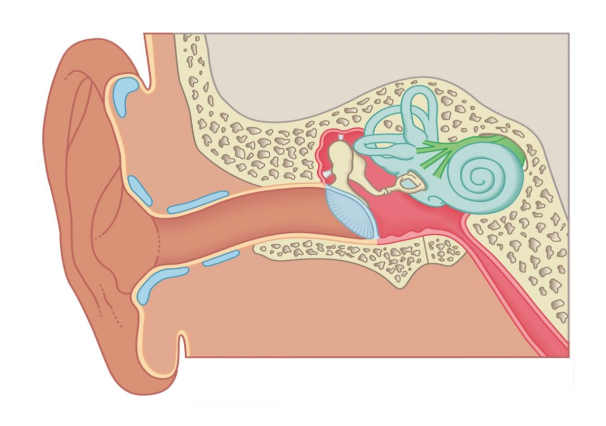


图3-3　人耳横截面示意图

声音频率对于听觉系统十分关键。人类听觉的敏感范围为20～20 000赫兹，但是对1000～4000赫兹的刺激频率最敏感，这个范围涵盖了人类日常生活起到关键作用的大部分信息，如婴孩的啼哭、引擎的轰鸣等。频率的变化不仅影响着听者对于所听内容的分辨，也包含着声音来源的信息，通过这些信息，我们可以知道发出声音的物体是什么。这是由于基底膜震动时，并非所有毛细胞都会随之震动，依据声音频率的不同，只有一些特定的毛细胞才会摆动。而发声物体具有的独特共振特性，可以使它与其他物体区别开来，即使是同样的音，在钢琴和古筝上被弹奏出来也会听起来完全不同。同样的道理，讲话时所发出的不同声音，也是通过改变声带的共振特性以及配合口腔、舌和嘴唇的运动完成的。日常生活中的自然声音是由复杂频率构成的，如此一来，特定的声音激活特定范围的毛细胞，从而在大脑中产生特定的听觉。

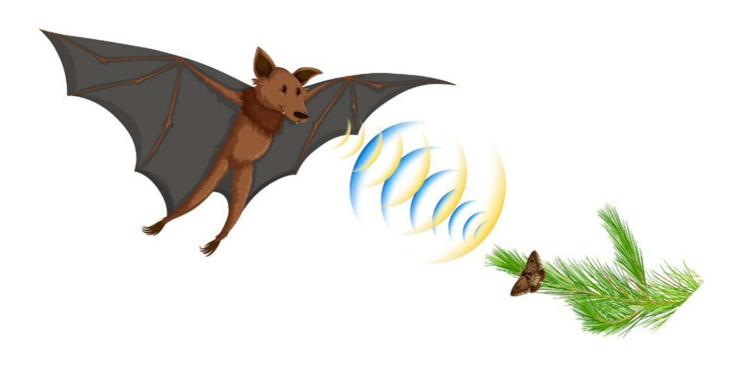


图3-4　蝙蝠回声定位

解决了如何听到和分辨声音这个问题之后，我们还需要了解声源的定位问题。你在生活中是否也有过这样的经历：在家里找不到手机时，叫人给手机打个电话便可以很轻易地根据铃声找到手机，这一过程我们称为定位。以听声定位的模范代表——蝙蝠为例，如图3-4所示，蝙蝠在飞行中发出高频声音，再由周围环境中的物体将这些声音反射回来，通过这些回声，蝙蝠的大脑建立了对周围环境和其中物体的声音图像，从而保证自己可以及时避开障碍物。虽然人类的听觉系统没有蝙蝠这么灵敏，但我们的大脑也可以对比进入两个耳朵的声音在强度和时间上的微小差异，以此在空间中对声源定位。

当一个声源发声时，声源到两耳的距离一般来说并不相等，因此到达两耳的声音也就不完全相同，而是具有一定的时间差和强度差。举个例子，如果一个声音从你的左侧传来，声波会先到达你的左耳，再传入右耳，这就形成了微小的时间差。同时，由于声音通过头颅时会造成一定的衰减，因此左耳所接收到的声音强度会高于右耳，从而造成强度差。这些信息会到达脑干的特殊位置，由其分析后将结果反馈给大脑，这样一来，我们便能对声音的来源进行精准定位。由此可见，听觉系统十分依赖整合双耳信息来实现定位。因此，在过马路时戴耳机听歌，哪怕只戴一只耳机都是很危险的行为。

以上情况是水平平面的声音定位机制。如果声源来自你的垂直平面，声源到两耳的时间差和强度差就没有太大变化，此时就无法通过刚才的2种方法进行垂直定位。那么我们是如何进行垂直定位呢？实际上，我们的耳郭具有不规则的褶皱（用手就可以摸到），这些褶皱

的作用便是对声音形成反射。当声源沿着垂直方向移动时，直接进入耳道和经耳郭褶皱反射后的声波组成的复合声，由于声源位置的高低而产生微小的不同，从而实现对声源的垂直定位。

## 五感的“记忆和情绪担当”——嗅觉与味觉

前面介绍的视觉与听觉能够使我们感受外界的物理信息，除此之外，我们还能感受环境中化学物质的刺激。比如，空气中有许多挥发性的小分子可以通过人呼吸或主动地吸气进入鼻腔，这就是嗅觉产生的基础；食物中的化学物质与舌头上的味觉感受器相接触，这是味觉产生的基础。嗅觉与味觉也是2种十分重要的感觉，这一点在动物身上尤为明显。对于一只昼伏夜出的小老鼠，它常需要通过气味来警惕天敌，一旦闻到细微的猫的气味，便赶紧藏好；也需要通过气味来寻找食物，并且判断食物是否无毒。味觉是对水溶性化学分子的感觉功能，能够作为嗅觉的补充来识别食物的性质，从而调节小老鼠的食欲并控制摄食量。可是为什么说嗅觉和味觉是五感中的“记忆和情绪担当”？不要着急，接下来就让我们一起来揭秘人体嗅觉和味觉的神奇之处吧！

嗅觉　带有气味的小分子可以通过3种方式进入鼻腔：第一种是随着正常呼吸或者我们主动去闻的过程中流入鼻腔；第二种是被动地流入鼻腔，因为鼻腔中的气压一般都比外界环境要低，气味分子会随着空气向气压较低的地方移动；第三种是通过鼻腔后嗅觉，进入口腔的气味分子也可以传到鼻腔内。气味分子进入鼻腔后就附着在位于鼻腔顶部黏膜中的嗅觉感受器上，即嗅细胞。嗅细胞是一种双极神经细胞，它的树突和轴突从其细胞体的反侧面延伸出来，当嗅细胞一端接收到气味分子时，信号就被传输到另一端的嗅觉信号处理第一中枢——嗅球。嗅球神经元（即嗅小体）离开嗅球后形成嗅神经，后者将信号传递给初级嗅皮质，也称次级嗅觉加工中心，它在判断是否有新气味出现时扮演着重要角色。

作为呼吸兼嗅觉器官，鼻子实际上有一个“秘密”。我们先来做一个小实验：把手指放在两个鼻孔的下方，然后正常呼吸，感受鼻子呼

出的气流。你有什么发现吗？

很多人并不知道，鼻子的左、右鼻道其实不是同时工作的，而是交替工作的，它们每30分钟到7小时轮换一次。当然，交替工作并不意味着休息的那个鼻道完全不参与呼吸，它只是在轮休时作为辅助鼻道而已。鼻子的这种轮班行为，在医学上被称为生理性鼻甲周期或鼻周期，这实际上是身体的一种自我保护机制。由于肺需要温暖湿润的空气，所以鼻子的一个重要工作便是温暖并加湿所有进入肺部的空气。如果不中断地让空气进入同一个鼻道，它就会变干甚至破裂，造成鼻出血或鼻科疾病，甚至影响嗅觉。

除此之外，左、右鼻道分别运行时，人的身体状态会有一些区别。当左鼻道运行时，身体器官的所有活动都会减缓下来，血压会降低。所以睡眠、犯困或情绪稳定时，多依赖左鼻道。与此同时，左鼻道还具有很强的判断力，辨别气味更准确。而当右鼻道运行时，整个身体的状态都处于活跃趋势，此时血压升高，每个器官都处于亢奋的状态。所以人在情绪波动时，多半用右鼻道呼吸。此外，用右鼻道闻东西时，往往对气味的印象会更深刻。

你是否有过被某种特殊的气味带回到很久以前记忆中的经历？影视剧中常常有这样的桥段：主角在外地偶然闻到某道菜的香气，便勾起了自己对家的记忆。为什么气味会与记忆紧密相关？一些科学家认为，嗅皮质与边缘皮质存在直接连接，而边缘皮质与记忆和情绪紧密相关。气味会比相关的视觉刺激更加稳定地激活边缘系统，从而触发记忆。

味觉　当食物刺激味觉细胞中的感受器，味觉系统的感觉转换就开始了。味觉细胞位于味蕾中，味蕾大部分都位于舌头上。基本味觉包括酸、甜、苦、咸和鲜，位于舌头不同区域的味蕾对于不同味道的敏感程度不尽相同，如图3-5所示，对甜、咸、酸、苦更为敏感的味蕾依次分布在舌尖、舌头两侧前半部分、舌头两侧后半部分和舌根。而“鲜”指吃牛排或其他蛋白质丰富的食物时所尝到的味道，并没有明显的敏感味蕾分布趋势。我们常说的“辣”其实并不属于味觉，而是痛觉的一种。你可能会好奇为什么有的人“无辣不欢”，而有的人却“滴辣不沾”呢？这是由于味蕾与疼痛纤维正好是相连的，味蕾越敏感的

人，痛觉感受器也越多，而且味觉能力受先天影响比较大，所以吃辣的爱好通常会遗传。不过，味蕾也会适应刺激，所以生活中也有很多人通过后天的饮食习惯，培养出了对辣的耐受度。




图3-5　味蕾分布示意图

每种味觉刺激都有不同的传导机制，能够转换不同的化学信号形式。味觉信息从味蕾传到初级味觉轴突，至脑干、丘脑，最后到达触觉及味觉皮层。初级嗅觉皮质与眶额皮质的次级加工区域相连接，人们所体验到的复杂味觉，就是由味觉细胞传递的信息经眶额皮质加工后整合得到的。除此之外，眶额皮质似乎也在加工摄入食物带来的愉悦感中起到了重要作用。

味觉与嗅觉常常被放在一起，不仅是因为这两种感觉都是化学感觉，还因为你能够尝到的味道在很大程度上都依赖于闻到的气味。正是由于刚才提到的鼻腔后嗅觉，进入口腔的食物分子也可以传到鼻腔内。实际上，我们对此早就有所体会——小时候抵触喝药，家长会告诉我们捏着鼻子喝会让药变得不那么苦；感冒鼻塞时我们往往因为尝不出饭菜的鲜美而食欲下降。这正是由于嗅觉信息获取受阻，导致我们无法“全方位立体化”地感受食物的味道。由此看来，我们形容一道菜好吃时往往称赞其“色香味俱全”也是有一定道理的。

通过上面的学习，现在我们可以回答这一小节最开始的那个问题了：无论是闻味思乡，还是抵触喝药，嗅味觉与记忆情绪间的密切联系使得它们当之无愧地被称为五感中的“记忆和情绪担当”。

## 五感的“‘弦’担当”——触觉

初见标题，你是否会感到疑惑：“‘弦’担当”是什么？物理学家认

为，弦理论有11维，包含10维空间和1维时间。类似地，我们标题里的“‘弦’担当”要表达的就是这样一种多维的概念。实际上，一切由皮肤-大脑完成的感觉都是触觉，它能够传递多维信息，包括触摸、压力、震动、温度、痛感以及四肢位置等。

皮肤是人体最大也是最早发育的感觉器官。当我们用皮肤触碰一个物体时，触觉系统可以告诉我们物体的形状、大小和表面结构等物理信息。皮肤下方包含了多种躯体感觉感受器，其中梅克尔小体探测一般的接触，迈斯纳小体探测轻微的接触，环层小体探测深层的压力，鲁菲尼小体探测温度。除此之外，疼痛由疼痛感受器或游离神经末梢探测——这些细胞有些有髓鞘，有些无髓鞘，它们的激活通常会使你立刻产生行动。在一些特殊情况下，比如手指被针扎或接触到高温的物体，有髓鞘的疼痛感受器会使你产生一个快速的缩手反应，这一反应一般由脊髓而非大脑控制，因此往往在你意识到疼痛之前，你的手就已经远离了“危险源”。而无髓鞘纤维则与最初刺痛后持续时间较长的更钝一些的疼痛有关，这是提醒你注意关照受到损伤的皮肤。

疼痛对身体来说是一种重要的警告功能，极端温度、伤口、发炎、腐蚀性试剂、中毒、外力等刺激都会触发该反应。当然，疼痛接收器也不会过度灵敏，而是需要刺激达到一定阈值时才会做出反应，其灵敏度受到组织内部的化学信使调节。当关节、牙齿或别的部位发炎时，人体组织会将体内环境酸性化，此时会产生大量的化学信使，用来降低我们的疼痛阈，这意味着我们对疼痛将变得更加敏感。因此，发烧的人往往感到浑身酸痛乏力，这便是身体为了避免遭受更多的伤害而强迫我们卧床休息，提醒我们要耐心静待痊愈。

你害怕打针时所带来的疼痛吗？在这里我可以教你一招：打针前可以提前按压将要接收针刺的部位，这在医学上被称为“加压麻醉”。其原理在于，按压皮肤所产生的刺激会叠加于痛感刺激上，从而相对减轻痛感。

有人说，带汗毛的皮肤可以传递情绪。1993年，瑞典的神经生理学家艾克·沃柏（Ake Vallbo）与同事首次在人类有汗毛的前臂皮肤上发现了特异性传递触觉情绪信息的神经纤维——CT（C-tactile）纤维。此后的研究表明，CT纤维仅被发现于有毛发生长的肌肤上，且对

接近人体皮肤温度的、速度介于每秒移动1～10厘米的触觉刺激具有十分强烈的反应。换句话说，CT纤维的功能与其他触觉神经纤维不同，它不仅传送触觉的物理信息，还在心理层面给予我们独特的感受。如此看来，“带汗毛的皮肤可以传递情绪”这一说法确实是有科学依据的。

实际上，触觉由2个系统组成，除了能够帮助我们解析触摸物体信息的感觉-辨识系统外；还有一个动机-情绪系统，帮助我们在人际互动中沟通情感。当我们伤心时，家人或朋友的拥抱能安慰我们的情绪；轻轻捏住小朋友婴儿肥的脸蛋，会自然而然地产生“小朋友真可爱”的想法；和喜欢的人并肩，手背不经意的触碰，会让自己的心里小鹿乱撞……这就是触觉的情感力量。研究表明，在悠久的生物进化过程中，人类已经发展出独立加工触觉情绪信息的神经网络。我们在接受轻抚或拥抱时获得的愉悦感，不仅是触摸动作的副产品，更是真正具有生存适应意义的。

## 运动控制

正如前文所描述的，知觉可以帮助我们探测、分析和估计环境变化，并做出恰当反应。“做出反应”往往涉及运动系统复杂的计划、协调和执行动作的能力。举一个简单的例子，现在正在阅读这段文字的你要伸右手拿起手边的水杯喝水，以你的躯干作为参照系，在这个过程中你的手从书的侧边离开，产生一个向右的位移直到水杯处，又产生一个向左上方的位移将水杯送到嘴边。手的运动状态从静止开始，经历向右的加速运动，向右的减速运动，静止，向左上的加速运动，向左上的减速运动，静止到嘴边，再由手腕发生旋转，使水能够自然流进你的口中。你的五指经历了从自然放松，到张开，再到握住水杯的过程。在这整个运动过程背后，是大脑一边在发出指令控制上肢的运动，一边根据视觉和触觉的反馈修正着运动控制，保证水不会洒到身上或书上。

一个简单而日常的动作背后竟然蕴藏着一系列如此复杂的反应，遑论那些更加复杂的运动了，如花滑运动员在冰面伴着音乐优美地舞

蹈、短跑运动员听到发令枪声后迅速起跑、钢琴家在琴键上娴熟的指法技巧……但总体来说，这些复杂运动控制可以抽象成一些简单的模型与类别，通过学习这些模型，可以帮助人们举一反三地理解运动控制的复杂过程。

## 运动是如何产生的？

根据运动的复杂性和受意识控制的程度，一般将运动分为3类：反射运动、随意运动和节律性运动。其中，反射运动是最简单的运动形式，膝跳反射便是最典型的代表，这类运动一般不受意识控制，运动强度与刺激大小相关。随意运动，顾名思义，一般是根据主观意愿行动，具有一定的目的性，其方向、轨迹、速度等均受意识控制，并且在过程中也可以“随意而行”，比如刚才提到的喝水动作。最后一类运动则是介于反射运动与随意运动之间的一种形式，你可以随意指使它开始或停止，但是在开始后不需要你有意识地“盯”着它，它可以自己一直重复下去，比如走路时左右胳膊的自然甩动，当你走路时将双手从口袋里拿出，自然地垂落，它们便会在你没有意识到的情况下前后摆动。顺带一提，这种走路方式才是符合生理的、正确的走路姿势。以上3种类型的运动便可将我们日常生活中所有的运动形式概括起来。

成年人一般有206块骨骼，它们组合成许许多多的活动关节，在600多块肌肉的作用下产生无数或简单或复杂、或快速或精细的动作。人体的任何运动都受到神经系统的调控，只不过简单的反射运动一般都由脊髓“处理”了，只有较为复杂的运动才需要忙碌的大脑“过问”。外部世界由上一节中介绍的感觉神经系统，将光、声、味、嗅、触等物理或化学能量转换，形成大脑细胞间可以交流理解的神经信号。复杂运动的计划、控制、学习、适应和掌握除了需要依靠这些感知觉的信息反馈，还常常受到注意力、主观动机和情绪等方面的影响。这些综合因素被大脑整合后产生对环境做出的复杂行为反应的指令，大脑运动系统再将其转换成一系列严格控制的肌肉收缩指令并下达，最终实现运动行为。这表明运动控制不仅和大脑的感觉系统有关，同时还与意识、学习、记忆等大脑的高级认知功能具有密切联

系。

刚才提到，简单的反射运动一般都由脊髓“处理”，实际上，在脊髓内部，有大量的协调控制某些运动的神经环路，特别是那些重复性运动的环路，这些运动被大脑的下行指令影响、执行和修饰。接下来，我们就来介绍运动系统的结构以及它们之间是如何互相联系的。

运动系统的“执行者”

首先我们来了解运动系统的基础结构（图3-6）。身体可以运动的部分称为效应器，除了那些离身体中线较远的远端效应器，如手、手臂、脚、腿等，还有离身体中线较近的效应器，如肩、肘、腰、颈等。上下颌、舌以及声道是发出声音的核心效应器，而眼睛则是视觉的效应器。

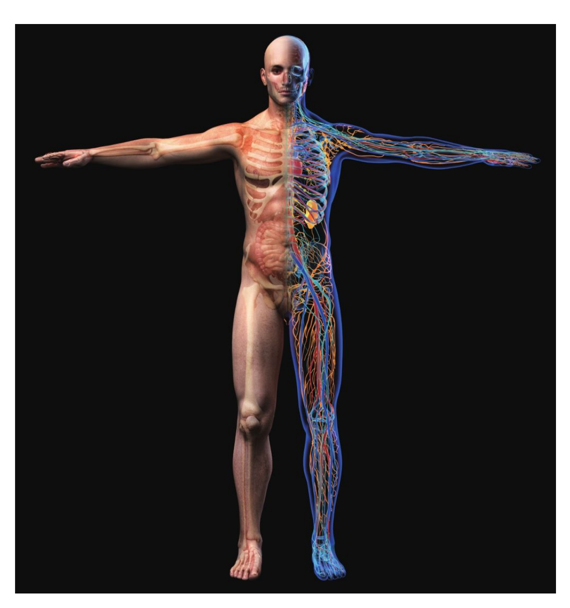


图3-6　人体肌肉、骨骼及神经的解剖示意图

各种形式的运动都由一个或一组控制肌肉状态变化的效应器产生。肌肉由弹性纤维组成，弹性纤维可以改变自身的长度和张力。这些纤维与骨骼在关节处相连，并通常组成拮抗的一对，即它们的作用结果相反，从而使效应器发生运动。例如，屈伸小臂时，肱二头肌和

肱三头肌就组成一对拮抗肌，肱二头肌收缩、肱三头肌舒张使肘关节弯曲，小臂屈起；反之，肱二头肌舒张、肱三头肌收缩则使肘关节伸展，小臂放下。

运动系统的“高层”们

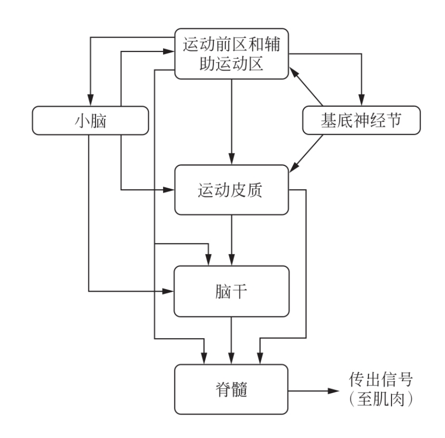



图3-7　运动系统的等级结构

正如前文所说，运动的脊髓控制只是最简单、基础的部分。如果将运动系统看作一个等级结构，位于最低层级的就是脊髓（图3-7），它提供了神经系统和肌肉的联系点，一些简单的反射运动也在这一水平进行控制。而位于最高层级的是大脑皮质的运动前区和联合区，这些区域负责运动计划，即确定运动的目标和达到目标的最佳运动策略。在小脑和基底神经节的帮助下，运动皮质和脑干将动作指令转化为运动战术，即肌肉收缩顺序、运动的空间和时间安排，以及如何使运动平滑而准确地达到预定目标，最后由最低层级——脊髓负责运动的执行。最高层级可能并不关心运动的细节，而是为低层级将运动指令转化为运动提供计划和指导。接下来，我们就来简单介绍一下这些运动系统中的高级层级。

运动系统的高级层级包括脑干、小脑、基底神经节以及运动皮质等（图3-7）。其中，脑干是维持个体生命最为重要的部位，包括心跳、呼吸、消化在内的一系列重要生理功能。小脑是一个忙碌的“中转站”，接受控制运动的各类信息又将这些信息整合转化进行传出的

工作，能够帮助保持运动的协调以及维持平衡。基底神经节是5个核团的总称，与小脑有一些相似之处，但其消息的输出主要是上行的，即通过丘脑投射至大脑皮质的运动区和额叶区域。大脑运动皮质可以直接或间接地控制脊髓神经元的活动，并制定运动计划。各个运动区域的协同工作最终使我们实现对运动的计划和执行。

通过对运动系统的高级层级进行简单了解，我们知道能够直接唤醒运动神经元的结构是脊髓，而脊髓可以接受来自2个上级的指示——大脑皮质和脑干。这些信息沿着2条主要的通路下行到脊髓，分别为外侧通路和内侧通路。外侧通路参与肢体远端肌肉装置的随意运动，比如喝水的例子，该通路受皮层直接控制。内侧通路参与身体姿势和行走运动，如我们在静止或运动状态下保持头部稳定，这一过程受脑干控制。

## 运动障碍患者的“身不由己”

到目前为止，我们了解了一些基本的运动生理学知识，也已经知晓神经通路对运动控制的重要作用。那么试想一下，如果与运动控制相关的神经系统出现问题，会发生什么严重的后果呢？如果人体遭遇了神经系统疾病、精神障碍或外伤等问题的侵扰，就会产生运动的兴奋或是抑制，甚至不能由意志所控制的现象，又称运动障碍。在本部分，我们将对常见的运动障碍疾病进行介绍。

皮质区域损伤

正如上一小节所说，大脑皮质可以直接或间接地控制脊髓神经元的活动。其中，运动皮质是负责掌控自主运动的区域，为熟练动作的产生提供最为重要的信号，同时，它也接受几乎所有参与运动控制皮质区域的输入，以及皮质下结构如基底神经节和小脑的信号输入，可见其作用不容小觑。

运动皮质的损伤通常会导致偏瘫。患者会因此失去受损伤脑区对侧身体的自主运动，这也强调了运动皮质对运动控制的重要作用。

偏瘫的病因多样复杂，任何导致大脑损伤的原因都可引起偏瘫，

其中脑血管病是最常见的原因，例如颅脑外伤、脑血管畸形等。在众多因素之中，脑卒中占据主导，高达90%以上的偏瘫由脑卒中造成。脑卒中疾病的形成是由于脑部血管突然的破裂或者血管阻塞，使血液不能流入大脑，最终导致脑组织损伤而造成惨剧。在2021年数据统计中，脑卒中已经成为我国成人致死、致残的首位病因。

当出现偏瘫症状后，患者通常会发现一侧的肢体完全不能运动。这并不是患者的意志或者意识产生了问题，他们往往竭尽全力但是依然无法移动自己的肢体。并且，患者从偏瘫中康复的概率很小。如果运动皮质遭受损坏，病人极少情况下能重新获得控制对侧肢体的能力。重新的运动存在可能，但只是在执行不需要独立控制和多关节协调的粗略运动时，例如，当一位患者的腿因偏瘫受到影响时，他或许可以再次走路，但是姿势却很难恢复到从前。

还有一些皮质损伤导致运动协调功能出现缺陷，这些缺陷并不能归结于偏瘫、肌肉问题所导致的无力、感觉缺失或动机缺乏。这类病症被称为失用症。从字面义来看，“失用”是指“没有动作”，在更广泛的意义上表示运动技能丧失。例如，一位双侧顶叶损伤的患者不能再继续做切鱼片的工作：她可以正确地将刀插在鱼的头部，并准备一刀砍下去——就像她过去已经做过上千次的那样，但是这时动作停止了，她表示：自己知道如何完成动作，但是却无法执行动作。此外，她还经常发现自己把装糖的碗放进了冰箱，或把咖啡壶放进了烤箱。她保留了肌肉运动的能力，但是不能把动作和相应的任务联系在一起，或识别出物体的正确用途。

神经学家还将失用症分为了两种子类型：意向运动性失用和观念性失用。意向运动性失用症的患者似乎有对预期动作的粗略理解，但是不能适当地执行动作。如果要求患者模仿如何梳头，他们可能会不断地用拳头敲击头部。观念性失用症则更加严重，患者不知道动作的目的，他们可能再也不能理解工具的正确使用方法。例如，一个患者用梳子来刷牙，这个动作表明他可以做出正确的姿势，但却使用了错误的工具。

皮质下区域：小脑和基底神经节损伤

大脑皮层以下所有的脑结构可以统称为皮质下，这一部分继续对皮质下结构的损伤展开说明。参与运动的皮质下结构主要包含小脑和基底神经节。其中小脑作为人体重要的运动调节中枢，起着对运动执行信息传入与传出的作用，能够保持身体运动的协调，维持身体平衡。醉酒者之所以会出现“无法走出直线”或是“行动不平衡”的问题，最主要的原因就是小脑细胞对酒精极其敏感。

基底神经节是另一个主要的皮质下运动结构，它包含着5个不同的核团。基底神经节任何一部分的损伤都会影响动作的协调性，不同的损伤位置造成的运动障碍形式也大有不同。受到不同损伤的影响，人类所具备的姿势的稳定性和运动间的精妙平衡被无情地打破，从而产生十分严重的后果。我们在这里主要介绍两种最为常见的病症：亨廷顿舞蹈症和帕金森病。

亨廷顿舞蹈症是一种退行性障碍，患者一般在40～50岁时开始出现临床症状。这种疾病在最初发作时并不明显，患者的精神状态逐渐改变：易激惹、神志不清、对日常活动失去兴趣。随着病程的不断发展，逐渐可以发现患者的运动出现异常，比如笨拙、平衡有问题，并且会不由自主地不停运动。这样的非自主运动，又称为舞蹈症，会逐渐支配正常的运动功能。患者的手臂、腿、躯干和头可能不断地运动且姿势扭曲。

事实上，在17世纪，科技尚未发达，那时亨廷顿舞蹈症不被大众熟知时，在至少两个大陆上，亨廷顿舞蹈症患者都会被指控使用巫术而遭到处决，因为病症发作时使他们看上去像是被邪恶的精神力量所控制一般。

亨廷顿舞蹈症造成的神经缺陷还不仅限于运动功能。随着运动问题不断恶化，患者还会发展出皮质下类型的痴呆症。患者可能会出现记忆缺陷，特别是在对新运动技能的学习上，并且在问题解决任务中很容易犯错误。

亨廷顿舞蹈症目前还无法被治愈。通过对患者的尸检发现，亨廷顿舞蹈症患者的大脑皮质和皮质下区域有大面积的病变，基底神经节的萎缩也明显可见，纹状体的细胞死亡率高达90%。

另一类广为人知的基底神经节疾病就是帕金森病。帕金森病分为阳性症状和阴性症状，分别指肌肉活动性的提高或降低。

阳性症状包括静止性震颤和肌肉强直。震颤是帕金森病的首发症状，大多会由一侧上肢的远端其他手指开始震颤，然后逐渐扩展到同侧下肢以及对侧的上下肢。这类的震颤频率大概在每秒4～8次，时而可以受到人为意识的控制，但却没有办法持久控制。在患者激动或者疲劳时震颤会尤其加重，睡眠时消失。

帕金森病的阴性症状是姿势和行进异常，运动功能减退及运动迟缓。帕金森病患者会失去正常的平衡功能。当他们坐着时，头可能不断向前下垂；而当病人站着时，重力作用会逐渐把人向前拉直到失去平衡。患者会出现运动功能减退，也就是自主运动的缺失或减少，他们的表现就像是牢牢定在一个姿势上不能改变一样，且在其试图发起一个新的运动时这个问题尤其显著。许多患者发明了一些小窍门以帮助克服运动功能减退。例如，一位患者拄着拐杖走路，不是因为需要它帮助保持平衡，而是因为它能够为其提供一个视觉目标以帮助其开始运动。当他想要走路的时候，他把拐杖放在右脚前，用脚踢它，促使自己克服惯性并跨出第一步。只要运动开始了，动作就显得正常了，尽管往往很迟缓。同样地，帕金森病患者可以用手拿物体，但整个动作进行的速度很慢。

## 记忆与注意

作为学生，你最怕在课本上看到的话是什么？我相信大部分同学都会回答是“背诵并默写全文”。我们总是感叹，要记得东西太多，脑力和时间却永远不够，考前熬夜复习的晚上，幻想拥有某种玄学或吃哆啦A梦的“记忆面包”能帮我们变成“量子速读小天才”。

记忆这件事总是让我们又爱又恨，不过先不用着急，虽然世界上没有“记忆面包”这种考试神器，但提高和改善记忆力却是有迹可循的。希望通过这节内容的学习，能够让你了解人类的记忆，并找到适合自己的记忆方法。

## 照相式记忆是真的吗

回想一下，你在生活中有没有遇到过这样的同学：他似乎有过目不忘的本领，背书就像吃了哆啦A梦的“记忆面包”一样轻轻松松。我就曾遇到过这样的学生，据她描述，在初中时她可以用一个午休的时间背下半本政治书上的内容，考试时，书本的内容就像照片一样印在她脑海里，甚至可以前后翻阅。听了她的描述，大家都感到十分羡慕，相信你们也是一样的。不过，这种照相式记忆真的存在吗？如果存在，我们能不能学会呢？想要知道问题的答案，就要先搞清楚记忆形成的过程——我们是怎么记住东西的？

总的来说，记忆包含了3个阶段：编码、储存和提取。你可以把大脑想象成一个图书馆，里面可以分成3个区域：大厅叫“瞬时记忆”，从外界接收的所有信息都聚集在这里等待下一步指示，有的记忆在这里待了一会儿就离开了，有的记忆则被大脑看中留了下来，从而进入另外2个区域——“短时记忆”馆和“长时记忆”馆。对这2个场馆的介绍将在下一小节中详细展开，现在我们先来了解被看重的信息是怎么留下来的。

你注意过图书馆藏书的书脊最下方所贴的标签吗？它通常由1位字母加4～6位不等的数字组成，可别小看了这个标签，有了它，熟悉图书馆的人就可以从琳琅满目的书中高效、准确地将目标定位，这就是编码的重要之处所在。类似地，大脑也会对将要储存的信息进行编码，给它们贴上独一无二的标签。不同的人看到同样的东西，会形成不同的编码，贴不同的标签，这些标签在认知科学中被称为“心理表征”。

心理表征包含了对信息的原始加工和心理加工。比如，考试取得了好成绩，父母奖励了你期待已久的礼物。当大脑编码这个信息的时候，会同时在脑海中涌现出这个礼物的形状、颜色、收到礼物时的喜悦心情、父母的表情，等等。除了这些简单直接的信息外，还有一些更深层次的信息也会出现：比如，你下定决心继续好好学习，或者要好好珍惜这个来之不易的礼物等，这些都是编码后的心理表征。

完成了编码，信息就可以正式入驻大脑这个图书馆了。接着便进

入记忆的第2阶段：储存。有的信息被送往“短时记忆”馆，有的信息被送往“长时记忆”馆。这个过程能不能很好地完成，与记忆的类型、个人的记忆能力以及大脑的健康情况等息息相关。

第3个阶段是提取。默写时绞尽脑汁地回忆书上句子的过程，就是大脑在进行记忆提取的过程。我们都有过这样的体验，有些知识点在考场上死活想不起来，越是紧张，大脑似乎就越空白，但是一出考场没多久就想起来了。这说明，记忆储存没问题，只是在考场上提取时出了问题。为什么提取过程会出问题呢？有两方面的原因：首先，可能是记忆的第1阶段——编码做得不够细致，就像整理文件时，文件说明写得越简单，事后查阅起来就越费劲；其次，可能由于过于紧张的情绪，人在紧张和焦虑时会分泌一种叫作皮质醇的激素，少量的皮质醇可以促进学习和提高注意力，但是大量的皮质醇会严重影响记忆的形成和提取。因此，如果在考场上遇到这种情况，不妨放下笔，做1分钟深呼吸，使自己冷静下来再继续答题。

了解完记忆的3个阶段，我们再回到最开始的那个问题：照相式记忆真的存在吗？在回答这个问题之前，我们再来做个小实验。现在，请仔细观察图3-8，10秒后再继续往下阅读。

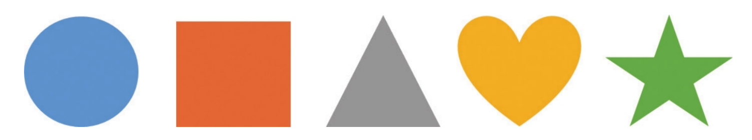


图3-8　五个图案

现在，请回答以下几个问题：这张图里有几个图案？分别是什么？它们的颜色呢？排列顺序呢？

如果你刚才有意识地记忆了这张图，那么你也许都可以回答上来，但假如你没有刻意去记，相信你凭“印象”也能回答上来1～2个。不过，刚才你回想的时候，脑海里的“印象”究竟是什么？

在专业术语中，这种即使并未刻意注意，但当你足够快地提取它时，会发现它仍在那里的记忆叫作瞬时记忆。由于瞬时记忆和视觉、听觉、味觉等紧密相关，因此又称感觉记忆。你能够在上课走神被老

师点起来重复最后一句话时有惊无险地回答问题，也是它的功劳。感觉记忆的容量很大，能包含许多信息，且由听觉保存的感觉记忆（又名声像记忆）会比由视觉保存的感觉记忆（又名图像记忆）保留的信息更多、时间更长。不过无论是图像记忆还是声像记忆，它存在的时间其实非常短暂，只能维持几百毫秒到几秒。当我们希望用感觉记忆来进行系统回忆的时候，往往已经来不及了。

如此看来，这种感觉记忆并不是我们要找的照相式记忆。

实际上，“照相式记忆”也有一个专业的名字，叫作遗觉像。与我们靠感觉记忆记住的模糊图像不同，据有这种能力的人描述，他们记忆中的图像与原始的图像一样生动，并且看起来好像是在“头脑之外”，而不是在“头脑之中”。不仅如此，它可以持续数分钟，甚至数日。遗觉像较多地出现在6~12岁的儿童之中，大约只有5%的儿童具有这种能力，且年龄越大，发生率越小。心理学家推测，这种能力的消失可能与儿童形式运算思维或者语言技能的发展有关，随着儿童不断发展语言和表达能力，遗觉像能力就会相应退化。

说到这里，我只能遗憾地告诉你，这种照相式的图像记忆实际上是不存在的，至少在你长大之后是这样。不过别灰心，正所谓“条条大路通罗马”，想要提高记忆能力，虽然照相式记忆不靠谱，但读一遍就能背诵的记忆能力还是存在的。至于这种记忆要怎样练成呢？这里先卖个关子，会在后续的学习中为你解答。

## 既视感是怎么回事

可能我们都听说甚至亲身经历过这种情况：走在熟悉的路上，突然感到莫名的危险，全神戒备地观察之后发现果然有可疑的人在附近游荡；老师曾经叮嘱过你，考试时遇到一道不确定的选择题，各种分析都用上了却仍不确定的时候，一定要相信自己的第一直觉；偷偷打游戏，即使将电脑提前关机散热，却还是被父母一下识破……

这些现象背后，就是生活中十分常见的——直觉在作祟，也有人把这种神奇的体验称为第六感，认为冥冥之中是宇宙的神秘力量在指引自己。

实际上，直觉的背后，涉及记忆里一个重要的概念：记忆的内隐作用和外显作用，简单理解就是记忆的“内化”与“思考”。比如，父母下班回家，进了家门之后，他们知道这是自己的家，而非走错到了隔壁。如果在这之前你偷偷打了游戏，动过鼠标、键盘、转椅，父母就能立刻察觉，虽然此时他们可能还不知道你动过什么东西，但就是感觉不对——可能是鼠标的位置被移动了，或者椅子的角度被转动了。这些细节，父母回家前可能并不记得原本的样子，但此时就是能知道不对劲。这就是记忆的内隐作用。如果你反问：这些东西和原来怎么不一样了？父母就会巡视房间并开始思考：是鼠标的位置不对？还是转椅？这个过程就是记忆的外显作用，需要做有意识的回忆。

其实，直觉就是一种经验法则，是无意识的内隐记忆起了作用，虽然并不一定准确，但总归是有迹可循的。但这时你可能会有疑问：为什么有时候我明明能确定自己之前从未去过某地，却还是产生了“我好像在哪里见过这个地方”的强烈直觉？据调查，70%的人都有过类似的经历，可能是某段对话、某种场景布局或者某种味道突然触发了某个开关，让你对过去未经历的事产生了浓浓的既视感。这到底是怎么回事呢？

在“照相式记忆是真的吗”中我们了解到，记忆的第1阶段是编码，大脑会给需要保存的信息贴上若干标签，今后只要经历符合这些标签，大脑就根据线索唤醒这段记忆。既视感可能就是由于当前场景与你的真实记忆或虚拟记忆相似而产生的。

对于前者，即“没有发生过，但与真实记忆相似”可以用以下原因解释：大脑中对相似度识别起作用的区域是顶叶皮层和海马回，这个系统“工作失误”时，会读取与现实相似度不到100%的记忆，也就是说，可能场景相似70%以上，大脑就把它看作当前场景的记忆来读取。

而对于后者，即“没有发生过，但与虚假记忆相似”的解释则更多一些：一种解释是由于你的关注点被打断，等注意再次回归时就会出现似曾相识的感觉。这是由于大脑负责处理接收到信息的左侧颞叶会收到2次相同的信息，这2次信息传递的路径不同：第1次是直接抵达，第2次则要先绕远到右侧颞叶再传回左脑，两者之间的延迟是毫

秒级的。但如果第2次到达时延迟稍长，大脑就会把这个迟到的信息标记为已处理过的，从而让你产生似曾相识的感觉。另一种解释是这个场景可能在梦中出现过。梦中大脑会产生很多场景，正常人每晚会做4～6个梦，虽然梦中场景的素材都是来源于你的记忆，但组合起来却可以是全新的记忆。然而，梦很少会留在外显记忆中，大部分都存在内隐记忆，醒来后，没有被记住的梦就好像沉入了意识的水中，很难被回忆，这也是很多人一觉醒来觉得自己没做梦的原因。这些“水下”的梦境记忆由于索引非常少，甚至可能只有几个模糊的间接索引，即使被某个场景或某句话所触发，让你产生似曾相识的感觉，也无法找到这段记忆的前因后果、具体信息。因为梦中的事件本来就是跳跃的，就算再努力想，你也想不到何时何地有过相似的感受。

## 大脑的内存与硬盘

在“照相式记忆是真的吗”的介绍中，我们曾提过大脑图书馆还有另外两个场馆：“短时记忆”馆和“长时记忆”馆。事实上，它们二者之间既有区别又有联系，用“内存”与“硬盘”或许可以更好地解释它们的关系。电脑的“内存”和“硬盘”相信大家并不陌生，前者是一个临时存放信息的小仓库，影响电脑的运行速度，内存越大，电脑能够快速调用的信息就越多，但是内存中的信息会随着电脑的关机而丢失；后者是一个大仓库，存放着电脑的所有信息，这些信息会在需要时被调出到内存中临时储存，但本身存在硬盘中的信息却不会因为电脑关机而清零。用一个简单的例子来比喻就是：你从口袋里掏瓜子吃时，硬盘就相当于口袋，其大小决定了你能装多少瓜子，而内存就是你的手，其大小决定了你一次能抓多少瓜子出来。

在介绍大脑的“内存”——短时记忆之前，你可以自己做个小测试：对于一串十几位的随机数字，只看一遍，你能记住多少？如果你能全部记住，那你的记忆能力算是相当不错的了。这个能记住的项目数量叫作“记忆广度”。其实关于短时记忆，我们在第1章中介绍认知心理学时就已经有所提及。短时记忆是一种持续时间非常短（仅比感觉记忆长几秒）的记忆形式，它是在感觉记忆的基础上，施加了注意成分所形成的。还记得米勒发表的那篇关于短时记忆广度的研究报告

吗？他在报告里提到，一般人的短时记忆广度在7个项目左右。你可能觉得，这也太少了，像刚才测试的那十几个随机数字，不说全部记下，但你好歹也能记个八九不离十。为什么会有这样的感觉？

这是因为，在进行短时记忆的时候，大脑往往使用了其他信息加工方式来扩展记忆广度。那么，你是如何做到的呢？除了第1章里提到过的“组块”法，复述也是一个常见的方法。比如，进入一个新的班级，你可能没法一下将所有同学的名字都记住，但是如果多发几次作业，每次都在心里复述同学们的名字，随着次数的增加，就能成功地记下所有人了。

要学习短时记忆，有一个重要的概念一定不能错过，那就是“工作记忆”。在有些文章中，工作记忆和短时记忆所表示的是同一个含义，不过在本书我们对这两个概念做以下区分：工作记忆代表一种容量有限的，在短时间内保存信息，并对这些信息进行处理的过程，因此被喻为“思维的画板”。其内容可以源于感觉记忆的感觉输入，也可以从长时记忆中提取获得。工作记忆概念的出现是为了扩展短时记忆的概念，除了在较短时间内的记忆外，工作记忆还有非常重要的功能，那就是信息加工和认知操作。简单来说，工作记忆仿佛一座桥梁，连接了感觉记忆与长时记忆，负责前者的写入和后者的读取，以及信息编码的储存。

如图3-9所示，工作记忆可以分为4个成分：中央执行系统、语音环路、视觉空间画板和情景缓冲区。事实上，它们4个不是平行关系，而是中央执行系统下辖另外3个成分。中央执行系统就像团队的领头人，负责帮你将注意聚焦在相关信息上，同时协调语音环路、视觉空间画板和情景缓冲区这3个团队成员对信息进行整合。

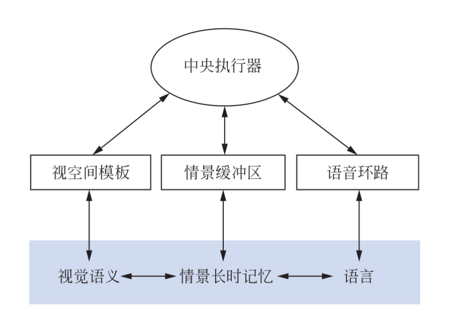



图3-9　工作记忆理论模型

你默读时头脑里会有声音吗？在语言记忆过程中，尤其是非母语的记忆过程中，我们常以声音的形式加工语音信息，默读时头脑里的声音就是语音环路在帮助我们提高短时记忆的一种表现形式。当然了，你会发现如果进行的是快速阅读，那么脑海里的声音就会因为跟不上眼睛的移动速度而逐渐消失，此时的阅读是直接从视觉到意义，跳过了语音通路。但是这种现象只能发生在我们熟悉文字内容的情况下，对于之前不了解的新字或非母语的阅读环境，默读是不可避免的。

视觉空间画板，顾名思义，就像一块画板，它能将所有的空间位置在大脑中展示出来。比如我让你回忆你在学校的座位，你的脑海里就会出现对应的画面。《神探夏洛克》中福尔摩斯的脑海里有一座记忆宫殿帮助他记忆，这个宫殿的打造就是利用了视觉空间画板。而情景缓冲区则类似于大脑的中转站，它一方面收集新信息，另一方面也从你的长时记忆中提取信息，同时把两方面的内容整合起来，变成我们可理解的信息。

如果说短时记忆是大脑的内存，那么长时记忆就当之无愧的是大脑的硬盘了，进入长时记忆的信息，可以在大脑里储存相当长的时间。长时记忆又可分为陈述性记忆和非陈述性记忆，实际上，就是我们之前提过的外显记忆和内隐记忆。前者可以通过我们有意识地回忆和再认而提取，比如你对学骑自行车那天（事件）的记忆，或者对自行车这一物体（事实）的记忆；后者则是一种无意识的记忆关联，比如学会骑车后，你坐上自行车就能自然而然地蹬脚踏板。

## 给大脑的硬盘升升级

本章最后，我们来解决大家一直关心的问题：怎样提高记忆能力？

不知道你有没有过这样的经历：从客厅走到卧室，刚一进去，忽然就忘了自己要去干什么；出门前惦记着一定要带某样东西，出门之后发现还是没带；明明锁了门，但是下楼后就开始忘记自己有没有锁门……这些情况让不少人非常担忧：自己是不是痴呆了？

不要担心，绝大部分人，尤其是像你这样的青少年，还远远到不了痴呆的临床症状，甚至离轻度认知障碍都还相差甚远。如果还是担心，可以上网搜索“简易智力状态检查量表”测试一下自己的认知情况，这是临床上筛查总体认知功能最常用的方法之一。测试完这个量表之后，相信你就可以完全放下心来了。

尽管如此，现代社会确实有越来越多的人或多或少地出现了记忆上的困扰。之所以强调现代社会，是因为在信息技术飞速发展的当下，无处不在的互联网确实在很大程度上影响了我们的记忆。

一项哈佛大学的研究发现，当人们知道自己所需要的信息可以在网上查到时，大脑就倾向于遗忘这些信息。由这项研究诞生了一个词——谷歌效应。人不可能记得住所有事，且大脑会自发地对获取的信息进行分类标记。显然，对于容易获取的信息，大脑就没什么必要将它们都储存起来。因此，人们以为被大脑储存下来的信息，其实大多被遗忘了，这种现象就是谷歌效应。可见，互联网的出现虽然令知识获取变得十分方便，却也改变了人们的学习和记忆方式。当我们认为某个信息可以通过搜索轻易获取时，对这个信息本身的记忆便减淡了，取而代之的是增强了对去哪里找到这个信息的记忆。

这项研究提示我们，移动互联网给我们生活带来的变化远比想象中的大，只是很多人没有意识到。诚然，信息技术的进步并不算是一件坏事，手机、计算机、互联网、云盘等，都变成了大脑的延展，只是我们的大脑在这个过程中自适应地调整了对信息的记忆和加工策略。简单地说，就是很多信息我们都“不往脑子里去了”。那么在这样

的时代背景下，我们该如何提升记忆能力、给大脑的硬盘维护升级呢？

在这里，有3条建议：

首先，多练习工作记忆。上一部分中，我们提过工作记忆可以看作是感觉记忆和长时记忆的桥梁，具有优秀的工作记忆能力往往代表可以在短时间内完成记忆的编码和储存，这有利于之后的提取。因此，你可以刻意地进行一些记忆训练，比如在规定时间内背下一篇课文或固定数量的单词，或者在路上无聊的时候试试能记住多少过往车辆的车牌号等等。

其次，采用多种编码形式。在本章开始的2个小节中，我们已经详细地介绍了编码对记忆提取的重要性。心理学家通常认为，编码的标签越接近语义，提取时就越有效。意思就是，根据某个词的使用含义去编码，对于这个词的记忆就更扎实。在此基础上延伸出来的理论就是，对于一个信息的编码方式越多，后期也就越容易提取。比如对于一个知识点，你上课认真听讲，课后及时完成作业，考前又认真复习，同时还能帮同学解答相关的问题，那么，你对这个知识点的掌握就比仅单纯反复背书的同学要好得多。

最后，主动地让“注意”参与到记忆过程中。还记得这一节内容的标题吗？——记忆与注意。虽然本章用大量的篇幅去介绍了什么是记忆，以及关于记忆的一些常见现象，但实际上注意对于有效记忆的重要性同样不可忽视。注意是人们留意一些东西的同时忽略另一些东西的能力，影响着我们如何分析感觉输入、编码加工输入的信息。认知科学研究中有这样一个有趣的理论：当我们和朋友在一个鸡尾酒会或某个喧闹场所谈话时，尽管周边的噪声很大，我们还是可以听清朋友所说的内容。这一现象为称为鸡尾酒会效应。它反映了在同一时间可以进入意识的信息量是有限的，我们不可能注意并同时处理所有作用于我们感觉器官的事物和刺激，大脑会帮我们选择性地注意一些重要的信息，而屏蔽其他事情。因此，在记忆的过程中，如果有太多的干扰因素分散了大脑对信息的注意，就很难对信息做出有效的编码。集中注意力，对信息进行更深入的编码，有利于短时记忆向长时记忆的转化，这样就能减少转身就忘的情况发生了。

可见，想要提高记忆能力，甚至做到“读一遍就能背诵”，其实并没有捷径可走，它需要你不断练习、不断努力以及全神贯注地投入。就像爱迪生说过的那样：“天才就是1%的灵感加上99%的汗水。”

## 小结

在本章中，我们首先介绍了视觉、听觉、嗅觉、味觉和触觉5种感觉，它们是人类认识世界的基础，让我们可以感知周围的环境。之后，进一步介绍了大脑的2个高级认知能力：运动控制、记忆与注意，阐明了这些功能相关的心理过程和神经机制，以及生活中与这些功能相关的常见事例。

# 4 “脑机接口”走进我们的生活还有多远

你是否幻想过用意念或者精神状态去操控机器，解救肢体功能障碍患者？随着不同工程科学的进步，为提高相关工作的安全性和有效性，降低系统的整体复杂性，减少执行任务所需的时间，同时增强系统能力，各种跨学科工程和人类协同集成设计的需求日益增加。这促进了机器控制的发展。机器控制是指从人体器官或神经系统中获取电子生物信号，然后从获得的信号中提取出特征，以此来确定人体的身体或精神状态和意图，最后，将不同的人类意图作为一种适当的控制命令转变成机器的物理动作。在这样的背景下，“脑机接口”逐渐走进我们的生活。

## 脑机接口技术

### 什么是脑机接口技术？

在过去的十几年中，脑机接口（brain-computer interface, BCI）成了一个非常重要的研究课题。通过解析大脑神经元放电信号得到分类指令，实现对外部设备（如脑控外骨骼、脑控轮椅等）的控制，脑机接口在医疗、军事、神经娱乐、认知训练、神经生物经济学等方面都有所应用。

2000年，第一次国际脑机接口技术会议将脑机接口定义为不依赖周围神经和神经的正常输出通路的通信系统。沃尔帕（Wolpaw）在综述中很有说服力地阐述了这一原则：“脑机接口将电生理信号从仅仅反映中枢神经系统活动转变为该活动的预期产物——对世界的信息和命令。它将反映大脑功能的信号转变为该功能的最终产物：像传统神经肌肉通道的输出一样，这种输出实现了人的意图。脑机接口用电生理信号以及将这些信号转换为动作的硬件和软件取代神经和肌肉以及它们产生的动作。”

脑机接口是一种基于计算机的系统，可实时采集、分析脑信号并将其转换为输出命令，涉及神经科学、机器学习、信号处理、机械工程、心理学等多个学科。通过脑机接口技术可以实现许多功能，如意念打字；士兵在战场上通过大脑远程操作机器人或无人机作战，可以减少人员伤亡；肢体功能障碍患者可以通过大脑控制物体移动等，比如轮椅行驶。类似以上这些情况的通过脑机接口技术最终实现脑和外部设备相互交流的方式称为脑-机器人交互。

在没有任何其他肌肉运动的情况下，脑机接口通过使用精神思维来控制外部装置的设备，从而在没有任何其他帮助的情况下提高残疾人的生活质量。作为一种新的意识输出和执行形式，用户必须有反馈才能提高它们执行电生理信号的性能。就像婴幼儿蹒跚学步、运动员或者舞蹈家完善自己的动作，用户的神经变化与输出必须与自身表现的反馈相匹配，才能调节优化整体表现，达到预期的目标。因此，大脑需要对行为反馈适应，脑机接口技术也应该能够进化到适应不断变化的用户大脑，以实现功能优化。这种双重适应要求用户和计算机都需要一定程度的训练和学习。计算机和实验对象的适应能力越强，所需的控制训练就越短。

脑机接口技术的实现共包括下述四个主要因素，如图4-1所示：

①信号采集。脑机接口系统所记录的大脑信号或信息的输入，然后将该信号进行数字化以便分析。

②信号处理。将原始大脑信息转换成有用的设备命令，这既包括特征提取，确定信号中有意义的变化，也包括特征转换，将信号变化转换为设备命令。

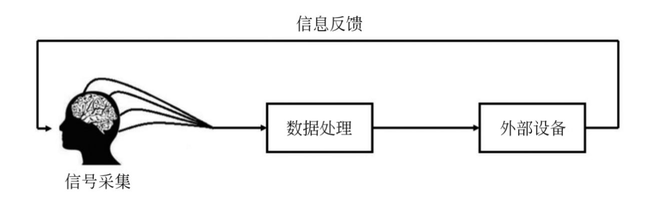



图4-1　脑机接口技术的实现过程

③设备输出。由脑机接口系统管理的命令或控制功能，这些输出可以是简单形式的基本文字处理和通信，也可以是更高级别的控制，例如驾驶轮椅或控制假肢。

④操作协议。系统开启和关闭的方式，这是指用户控制系统如何运行的方式，包括打开或关闭系统，提供何种反馈以及反馈的速度、系统执行命令的速度，以及在各种设备输出之间切换。

## 脑机接口技术的前世今生

1929年，人类脑电波的发现者——汉斯记录了人类脑电波活动后，利用思想控制机器便从虚构的想象逐渐进入科学探索阶段。1934年，科学家艾格和马修开发了获取脑电波信号的脑电设备。同年，生理学家费舍和鲁文发现脑电波信号中的尖峰信号，第一个脑电图实验室在波士顿建立。1940年，来自西北大学的生物物理学教授富兰克林开发一种脑电图模型来检测动作时的脑电波。1950年，威廉·瓦特发明脑电地形图，以描绘头皮周围的电活动。这项技术为神经学家和研究人员识别大脑信号记录提供了途径。脑机接口这个词可以追溯到20世纪70年代杰克斯设计的一个使用视觉诱发电位的脑机接口系统。从那时起，计算机技术、机器学习和神经科学的进步使得各种各样的脑机接口系统得以发展，脑机接口技术的探索逐渐火热。1969年，华盛顿大学医学院利用猴子进行脑电生物反馈的研究。1980年，施密特利用微电极将长期侵入性脑机接口连接系统与中枢神经系统连接以控制外部设备。20世纪90年代，杜克大学的尼可莱里斯完成对老鼠运动脑电波的初步研究，从脑电波中收集的信号被转换成思维来控制机器人。在2000年，尼可拉斯成功地在一只夜猴身上实现了侵入性脑机接口技术，它通过操作操纵杆来重建手臂运动以获取食物。经过升级后，猴子能够通过视觉反馈控制机器人手臂的运动，通过视频屏幕上移动的光标来抓住物体。2014年，科学家通过脑电图与经颅磁刺激技术实现无创的脑对脑直接交流。2019年，科学家利用人工智能将脑信号转化为语音并进行播放。图4-2中展示了脑机接口技术的发展史及标志性事件。











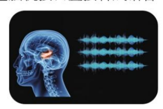




图4-2　脑机接口技术发展史及标志性事件

近些年，各国纷纷将脑机接口纳入重点研究的方向。2012年，加拿大创造了具备简单认知能力的虚拟大脑；2013年，美国政府正式提出“推进创新神经技术脑研究计划”，同一年欧盟委员会宣布“人脑工程”为欧盟未来10年的“新兴旗舰项目”；2014年，美国重点资助了9个大脑领域的研究，包括著名的“DAPPA”大脑计划、“阿凡达”计划；2015年，加州理工学院的科研团队通过读取病人手部运动相关脑区的神经活动，成功帮助一位瘫痪10年的高位截瘫病人通过意念控制机械手臂完成喝水等较为精细的任务；2016年，荷兰乌特勒支大学的研究团队通过脑机交互技术，使一位因渐冻症而失去运动能力及眼动能力的患者通过意念实现在电脑上打字，准确率达到95%。

中国在该领域的起步相对较晚，不过国内高校在脑机交互技术研发方面非常踊跃，清华大学、天津大学、浙江大学、北京理工大学、华南理工大学等高校在脑机接口的研究中处于领先地位。近些年，我国也逐步增加了在脑科学领域的投入，从2010年的3.48亿元，2013年的近5亿元，到近几年数十亿元的资金投入，充分体现了中国在该领域取得突破的决心。

近年来，脑机接口研究主要集中于运动想象和稳态视觉诱发电位（运动想象和稳态视觉诱发电位是实现脑机接口系统的两种不同方案）。国内的很多研究小组在稳态视觉诱发电位-脑机接口领域取得了不错的成绩，比如清华大学的生物工程研究团队在稳态视觉诱发电位的信息传输率研究上处于世界前列；华南理工大学李远清教授带领团队研究混合脑机接口系统在稳定性和可靠性上取得良好的成绩；2016

年10月，由天津大学神经工程团队负责设计研发的在轨脑-机交互及脑力负荷、视功能等神经工效测试系统随着“天宫二号”进入太空，完成国内首次太空脑机交互实验。2018年，华南理工大学的研究人员对非植入式脑控双机械臂进行研究，实现脑控双机械臂的运动，并且优化实现了对机械臂稳定性和协调力的控制。与此同时，从2000年开始举办的国际脑机接口竞赛，大大推动了脑机接口技术的研究。北京也举办了两届脑机接口比赛，此比赛要求参赛团队全方位完善脑机接口系统，从系统优化到性能评估，从离线到在线竞赛方式，大大提高了脑机接口技术的研究水平，部分成果在国际高影响力期刊上发表。

## 医疗领域中的脑机接口

### 让“假如给我三天光明和声音”成为现实

大家一定对海伦·凯勒的故事不陌生。这位美国女作家出生于19世纪，幼年因病失去视觉和听觉，但即使生活在一个没有光和声音的世界里，她仍然刻苦学习和写作，《假如给我三天光明》鼓舞了一代又一代人。然而，如果海伦·凯勒出生在当今时代，“三天光明”甚至“三天听觉”都有可能在脑机接口技术的支持下成为现实。

相信大家都对《黑客帝国》系列电影记忆犹新。在《黑客帝国》中，现实世界的人类通过在身体里插入连接器的方式实现和“母体”世界的连接，人类的意识可以通过这个“接口”进入电脑，所有的知识都能够以数据的形式下载到大脑里，每个人都可以在短时间内迅速学会功夫、甚至成为一名全能型“学霸”。这便是典型的“侵入式”脑机接口。

随着脑科学研究的深入，以及脑机接口技术的高速发展，脑机接口技术正逐步从科幻世界渗透到现实生活。利用大脑信号直接操控外部机器，以及利用外部信号刺激绕过神经系统，直接对人的大脑产生刺激等电影中才能看到的场景，已经逐步在现实世界实现。至此，不少科学家做出了与《黑客帝国》男主人公尼奥相同的抉择，在“蓝红药丸”中选择了红色药丸，致力于利用脑机接口技术攻克如今医疗领

域面临的诸多难题。

通过第3章的学习，我们知道人眼是人体工程学上的一个奇迹，是人类最重要的感官之一，也是我们的心灵窗户。人体所有感官的受体有70%位于眼睛，大脑皮层中有40%被认为与视觉信息处理的某些方面有关联。每个人都希望自己的眼睛明亮又健康，能够清楚地看到这个美丽的世界，感受一切色彩与光明。然而，当今全球仍有5000多万盲人，至少有22亿人受到不同形式的视力障碍，2.85亿人（该数据来自于2021年欧洲议会残疾人论坛）视力受损，对于他们来说，恢复正常视力，甚至重见光明都是一个遥不可及的梦。然而在脑机接口技术的存在的支持下，一种不需要视觉刺激系统直接参与，而是通过将光学信息直接发送到大脑的视觉皮层，从而让大脑直接获得基本视觉的方法已成为可能。那么这种利用脑机接口技术实现视觉的“人工眼球”（如图4-3所示）是怎么工作的呢？

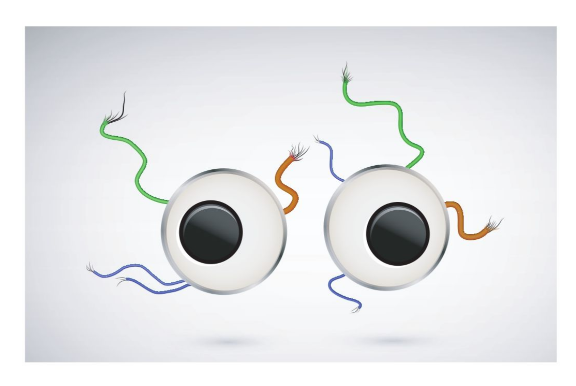


图4-3　“人工眼球”概念图

首先我们再来复习一遍视觉产生的机制。我们知道光是人类视觉刺激的关键。光本质上是一种电磁辐射，可以通过刺激视网膜从而产生视觉。电磁波按波长可以分为无线电波、红外线、可见光、紫外线、X光和伽马射线，可见光又根据不同波长分为红色、橙色、黄色、绿色、青色、蓝色、紫色等颜色。人眼只能对其中很小范围，即大致为380～740纳米的波长，产生视觉。人眼接收的是物体反射的光，我们看到的世界是五颜六色的，这是由于物体的可见颜色取决于其吸收或反射的光的波长，例如一般植物的叶子反射绿色的波段，吸

收其他颜色波长的波段。

人的眼睛本质上是一个复杂的光学感应器，由角膜、瞳孔、晶状体、玻璃体、视网膜、感光细胞等结构组成，功能上与照相机比较类似。照相机成像的原理是，光通过一系列光学元件后，完成折射和聚焦，穿过光圈孔到达成像平面，从而形成图像。人眼的各种结构实现类似的基本功能：角膜和晶状体实现聚焦功能，虹膜类似光圈控制装置，可以控制光通量，光通过这一系列结构后落在几乎透明的视网膜上，直至其最深的一层色素上皮层，然后反射回布满感光细胞的相邻层。感光细胞根据形状的不同分为视杆和视锥，从色素上皮层反射的光刺激感光细胞后，改变了其电性能并释放神经传送体刺激相邻的神经元，从而使神经脉冲在细胞间传递，传至神经节细胞的轴突后通过视神经和视觉盲点传至大脑的视觉皮层，最终形成视觉。

对于绝大多数盲人来说，视觉受损的主要原因是眼睛或视神经受损，而大脑皮层的视觉中枢可以正常工作。因此，为了更好地解决视觉受损的医学难题，科学家们致力于开发一种“人工眼球”设备，可以绕过受损的眼睛或视神经，直接将外界图像信息传输到大脑皮层，从而形成视觉。

基于脑机接口的人工眼球不再依靠感光细胞、视觉神经元，也不利用视觉细胞光信号-电信号的转换过程，而是利用体外处理器将图像信息进行人工处理与编码，将光信号直接转换为电信号，再通过插入的微电极阵列传导到大脑皮层的视觉中枢进行刺激，形成视觉。具体来说，“人工眼球”系统在大脑皮层的视觉中枢上植入微电极阵列，再将植入物与外界的图像采集设备、图像处理设备配对，其中图像采集设备为一副中央安装了摄像机的眼镜，用于采集眼前的图像信息，图像处理设备用于光信号和电信号之间的转换，同时将信号传输到大脑皮层。人工眼球工作时，摄像机捕捉进入使用者视野的图像，并将这些图像信息发送到计算机，计算机对其进行人工处理与编码，将其转换成电信号，并传输到微电极阵列。电极对视觉神经系统进行刺激，使盲人形成视觉。

早在1996年，来自美国犹他州的犹他大学的研究人员就成功开发出了这种“人工眼球”。在他们的实验中，研究人员将排列着100个长

度为1.5毫米、面积为12.96平方毫米的针状金属薄片电极植入失明患者大脑皮层，成功让其产生了“光幻觉”，患者可以描述研究人员预测的颜色，并且随着光斑的位置转动眼球。然而，为了帮助盲人患者形成视觉，仅仅呈现光幻觉是远远不够的。我们知道电子屏幕是由一个个像素点构成的，因此有人提出，如果电刺激视觉皮层产生的小光点的视觉感知能够结合成连贯形式，类似电子屏幕上的像素，是不是就可以在盲人患者的大脑皮层形成一幅完整的图像了呢？

为了让“人工眼球”更好地帮助盲人看到世界，2020年5月，国际顶级期刊《细胞》上发表了一项来自美国贝勒医学院丹尼尔教授团队的研究成果，该团队通过动态电流电极刺激大脑皮层，成功在失明患者脑海中呈现了指定的图像。丹尼尔教授表示：“当我们使用电刺激在患者大脑上直接追踪字母时，他们能够‘看到’预期的字母形状，并正确识别出不同的字母。他们把这些字母描述成发光的斑点或线条，就像正常人看到天空中出现的字母一样。”该团队对传统的电极进行改进，结合电流转向和动态刺激，通过对电流进行精准控制，依次激活不同的电极，实现字母或图片轮廓的绘制。如此看来，帮助盲人“看到”更复杂的信息，实现他们看清世界的梦想指日可待。

除了视觉外，听觉也是人类感知世界的一个重要渠道，是人类与外界沟通最重要的手段之一。然而，听障人士也是一个庞大的群体，全球大约有4.66亿人和海伦·凯勒一样患有残疾性听力损失，其中3400万人是儿童，且这一数字仍在上涨。他们因为听觉障碍影响了与外界的交往及生活质量，因此利用脑机接口技术恢复听觉也是我国医疗技术发展的一个重要方向。“人工耳蜗”是我们最早开发并成功应用的脑机接口技术之一，它可以为患有严重感音神经性耳聋且传统助听器无效的人提供人工听觉。“人工耳蜗”又是怎么工作的呢？

在了解人工耳蜗（如图4-4所示）的工作原理之前，我们再来复习一下听觉的产生过程。听觉的产生比较复杂，一般是声音通过空气传导。我们平常看见的“耳朵”，其实是外耳的耳郭部分。耳郭负责收集外界的声波，使其顺利汇聚入外耳道，通过耳道引起鼓膜的振动，进而引起和鼓膜衔接的中耳的听小骨的振动。听小骨的振动将声波转化为压力波，传递给耳蜗。耳蜗之所以叫耳蜗，是因为形似蜗牛壳，

但耳蜗里充满液体和毛细胞，液体的扰动会造成毛细胞弯曲，毛细胞就会制造神经信号，信号通过内耳神经，传导入大脑皮层中的听觉中枢，从而引起听觉。

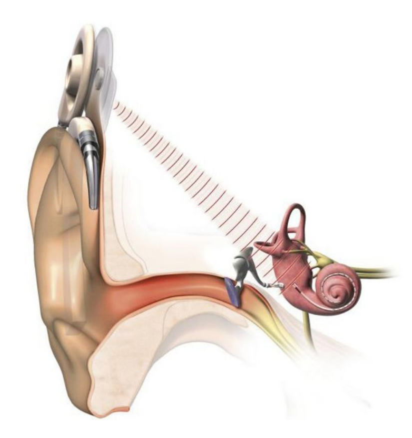


图4-4　人工耳蜗结构图

人工耳蜗不再依靠外耳、中耳的传导和放大功能，也不通过从声信号到电信号的转换过程，而是靠体外处理器将声音转为电信号并直接刺激听神经，再传导到听觉中枢产生听觉。人工耳蜗主要包括植入体和体外机两个部分。体外机负责接收声音，并将其转换成按语言信息编码的电信号。植入体通过植入式手术放置于耳后的颞骨表面，参考电极植入骨膜下，工作电极植入耳蜗内，植入体接收到电信号后，电脉冲通过电极通道序列刺激神经，从而产生听觉。

看来，如果海伦·凯勒出生于21世纪，在脑机接口技术的支持下，她不仅能实现“假如给我光明”，还能够重获声音，拥抱这个五彩斑斓、鸟语花香的世界。未来脑机接口领域还将创造怎样的医学奇迹呢？让我们一起拭目以待！

## 帮助肢体残障患者重新“动起来”

据中华人民共和国民政部统计，截至2021年，我国60岁以上的老年人数已达到2.67亿，目前我国老龄化趋势非常严峻。随着各类残疾人和长期卧床的老年人数不断增加，如何助老助残已经成为一个十

分严重的社会问题。随着人-机器人交互技术的发展，机器人在助老助残、医疗康复等领域扮演着日益重要的角色。机器人有望为老年人和残疾人提供居家养老、残障人士主动护理和神经系统疾病患者主动康复等全方位的服务，为提高生活质量、保证我国社会的稳定发展发挥重要作用。

近些年来，脑机智能技术的快速发展，为人与外部世界提供了一种全新的沟通交流方式。例如，脑机接口可以使失去活动能力的患者恢复其语言功能、行为表现等，如实现语言功能丧失患者的外界交流；辅助四肢完全丧失功能的残疾患者在无人照看的情况下操作轮椅；帮助渐冻症、脑卒中等患者提高生活质量与生存能力。研究人员尝试使用脑机智能技术去控制机械臂，外骨骼、控制导航医疗机器人的进行运动，为老年人提供了一种辅助生活的便捷方式。

但是这方面的研究仍存在以下问题：

①大多数脑机智能系统人机交互做得好，用户体验不好，用户无法实时了解被控外设的位置。很多研究将导航机器人的控制与稳态视觉诱发电位刺激之间的关系割裂开来。

②传统控制方法效率低下，用户通过控制机器人的前后左右使其缓慢移动到目的地。由于机器人控制算法的局限性加上脑机接口系统的延迟，机器人的控制变得更加困难。

针对上述问题，北京理工大学的脑机智能与神经工程实验室首先搭建了基于脑机接口的脑控智能机器人系统，为患者的术后康复训练提供了平台。首先，为了提高脑电采集设备的精度，该系统将伪迹子空间滤波算法植入到脑电采集设备中；其次，为了增强人与机器的交互以及对环境的感知，设计了基于机器视觉的动态虚拟现实和稳态视觉诱发电位相结合的范式，在真实环境下对物体进行识别与追踪，并使用闪烁块对物体进行标记；基于虚拟现实-稳态视觉诱发电位的在线脑机接口系统，开发脑电预处理算法、特征提取和分类算法，并进行验证，为脑控机器人的应用提供了很好的技术支持；最后，设计了基于人机协调控制的多自由度脑控机器人样机，对各组成模块作了详细地设计和验证，以满足实时的控制任务需求。北京理工大学的脑机智

能与神经工程实验室提出的智能机器人系统平台架构如图4-5所示。



脑控轮椅

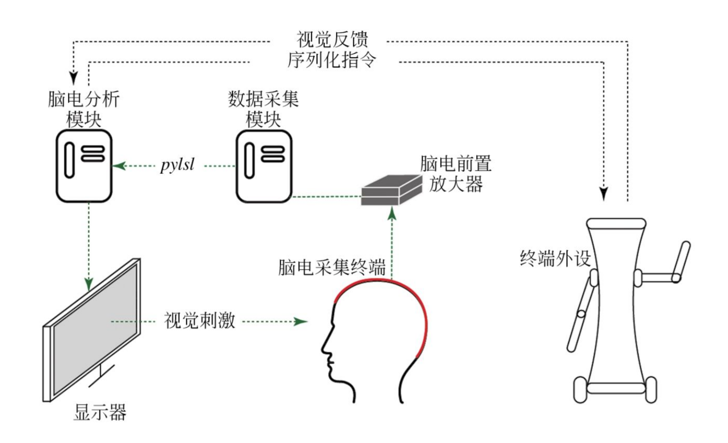



图4-5　智能机器人系统平台架构

北京理工大学的脑机智能与神经工程实验室提出的智能机器人系统平台主要包含以下两项核心技术：

①设计了基于机器视觉的动态虚拟现实和稳态视觉诱发电位相结合的范式。传统的稳态视觉诱发电位脑机接口控制系统无法与现实世界进行交互，长时间的闪烁刺激容易引起人类视觉方面的疲惫，影响识别精度。为了增强人与机器的交互以及对环境的感知，该系统设计了增强现实和稳态视觉诱发电位结合的范式，在真实环境下对物体进行识别与追踪，并将闪烁块对物体进行标记。



脑控机器人



图4-6　美国国防高级研究计划局意图实现通过“意念控制”远程操控“机甲战士”

②搭建了基于人机协调控制的多自由度脑控机器人设备样机，如图4-6所示，将机器智能与人类智能相结合，脑控作为第一控制指令，机器人在人类智能的决策下执行相应的智能化作业，从而满足复杂的作业任务需求。机器人领域正朝着智能化的方向发展，机器人可以通过传感器感知周围的环境和自身的状态，并能进行分析判断，然后执行相应的行动。虽然现实生活中机器人表现出来的智能化水平已经令人惊叹，但离它理想的状态还存在一定差距。在某些方面，比如环境适应能力、自主控制能力和环境感知能力等，机器智能是无法超越人类的。机器智能比较擅长计算机运算，在算法的可推广性方面更具优势，同时也更擅长于长时间运算，但是在逻辑思考能力方面不如人类智能。比如，在避障方面，机器智能经过多年的发展也只能实现简单场合的避障行为，在复杂场合下远远低于人类智能。深度学习技术快速发展，在图像处理、自然语言处理和视频处理等方面表现出了卓越的性能，但是在逻辑思考、危险预判方面还有着非常大的缺陷。虽然人类学习需要消耗比较长的时间，但是人类一旦掌握基础知识之后，便能够进行相关的逻辑思考、类比推流，人类智能的这种优势刚好弥补了机器智能的不足之处。所以，人类智能与机器智能两者相互依存，不可分割，两者是互补的状态。

目前我们国家肢体残疾患者数量呈逐年上升趋势，北京理工大学的脑机智能与神经工程实验室所研发的脑控机器人设备即主要面向肢

体功能缺失的患者。同时随着看护成本的增加，许多家庭也急需一台辅助设备帮助肢体残疾患者完成日常生活。

## 军事领域中的脑机接口

### 科技强军中的“脑控”技术

你是否曾幻想过通过“意念控制”像阿凡达男主一样骑着“坐骑”驰骋在天空？你是否也希望像阿凡达男主一样拥有第二分身？如果我们也能够通过“意念控制”体验远在千里之外的美景，那感觉将会多么美妙！我们还可以更进一步地思考，影片中的“意念控制”技术如果能够应用于战争中，将会极大地提高现代战争的效率，降低战争的伤亡。

科技兑换想象，科技不止一次将影视剧中的幻想带到了现实中。幻想是科学的来源，幻想总是会走在科学前面。更为确切地讲，所有科技的进步都源自于人类的梦想，如果没有梦想，人类就不会有研究的动力。现代科技从来不会让我们失望，《阿凡达》中所涉及的“意念控制”已经不再是导演卡梅隆的幻想。其实，“意念控制”就是基于脑机接口技术实现的。脑机接口技术可以让使用者拥有感知外部世界并通过“意念控制”操控物体的能力。《阿凡达》中男主所佩戴的设备就是脑机接口设备，科研人员通过计算机技术解读男主的脑电波信息，将男主的运动意图转化相应动作来驱动阿凡达的身体。

其实，美国国防高级研究计划局早在60年前就已经开始了对“意念控制”的研究。说起美国国防高级研究计划局，就不得不谈一谈这个机构所成立的背景以及所肩负的使命。美国国防高级研究计划局是美国国防部属下的一个行政机构，负责研发用于军事用途的高新科技。美国国防高级研究计划局成立于1958年，当时正值美苏冷战时期，双方积极展开备战，希望能够在军事领域占据领先地位。也就是在这个时间，苏联先于美国在1957年10月4日发射了“斯普特尼克1号”卫星，这使美国感到了前所未有的危机。于是美国国防高级研究计划局顺势而生，肩负着保持美国军事科技较其他的潜在敌人更为尖端的使命，大力从事超前的国防科技研发。如同美国国防高级研究计

划局的自述：“从1958年创立起，美国国防高级研究计划局的最初使命，是为了防止如同‘斯普特尼克’发射的科技突破，这标志着苏联在太空领域打败了美国。这个使命宣言也随着时代而演进。美国国防高级研究计划局的任务仍然是防止美国遭受科技突破的同时，也针对我们的敌人创造科技突破。”

脑机接口技术是美国军方美国国防高级研究计划局的一个重要研究分支。由于脑机接接口技术能够实现人脑对于武器装备最为直接的控制，能够赋予现代武器装备高度智能化的性能，该技术受到美国军方的高度重视。美国国防高级研究计划局在脑机接口领域投入巨大，通过向一些美国本土的研究机构资助研究经费开展相关研究。2004年美国国防高级研究计划局投入2400万美元用于资助美国杜克大学神经工程研究中心等6个实验室开展意念控制机器人、脑听器、心灵及生理响应系统、无线电催眠发生器等多项“脑机接口”技术产品的研发工作。其中意念控制机器人项目旨在打造可由士兵“意念遥控”的机甲战士，从而可以实现在战场上完成人类不可能完成的任务。2013年美国国防高级研究计划局资助了一项名为“阿凡达”的科学研究项目，目的是在未来使士兵能够通过“意念控制”远程操控“机甲战士”（如图4-6所示），从而代替士兵完成各种战斗任务，这正是从电影《阿凡达》中得到的启发。在2015年，美国国防高级研究计划局立项资助了一项战斗机相关的研究，目的在于赋予战斗机飞行员同时操控多架飞机和无人机能力。直到2018年9月，美国国防高级研究计划局宣称：“借助脑机接口技术和辅助决策系统，战斗机飞行员已能同时操控3架不同类型的飞机。”美国空军已经能够利用脑机接口技术提高战斗机飞行员的快速反应能力。

2021年7月，美国发布了《脑机接口在美军事中的应用及建议》，评估了脑机接口技术在军事领域的潜在应用，并且提出了未来可能面临的风险及相应的解决措施。随着科技的发展，世界主要科技强国纷纷意识到脑机接口对科技强国的助推作用，纷纷开始开展相关研究占据主动权。同样，脑机接口技术也是中华人民共和国国民经济和社会发展第十四个五年规划中重点发展的一项关键技术。

未来，脑机接口在军事领域的应用主要可分为以下3个方面：①

仿脑技术：武器的“智能”可能接近人类。仿脑技术借鉴人脑的运行机制，开发出具有类脑信息处理机制、仿生认知活动和智能化行动能力的高智能机器人。它的整体智力水平可能接近人类。②脑控技术：利用思想控制对抗武器将成为现实。脑控制技术借助脑机接口等建立人脑与智能设备之间的连接，基于检测到的脑电波信息编译计算机语言，实现人脑与智能装备之间的双向信息传输，实现智能装备的直接控制，减少甚至取代人体肢体运动，最终实现武器装备作战灵活性、敏捷性和效率的飞跃。在脑控技术的支持下，思想战争成为可能，大规模机器人将填充未来战场，人类在复杂战场环境中的生理极限将被打破，人类将可以成为“运筹帷幄之中，决胜千里之外”的决策者。③控脑技术：让敌人受制于己方意志。控脑技术利用外部干预技术干扰甚至控制人们的神经活动和思维能力，导致对方精神失常或幻觉，迫使对方在不知不觉中做出违背自身意愿的决定。控脑技术的关键是监控、收集和干扰大脑思维活动。控脑技术的基本原理是致幻剂效应，即大脑受到外部信号干扰后，被控制方根据信号的意图做出决定并采取相应的行动。

## 航天员或许也可以“躺平”

近年来，我国的航天事业飞速发展，我们能够直接看到宇航员在船舱中的实时状况。大家都知道，太空中宇航员会一直处于失重的状态，行动是非常不方便的。尤其是进行一些需要出舱的操作时，笨重的太空服会让一些地球上能够轻松完成的动作变得更难更慢。事实上，宇航员们都是经过相当强度的体能训练的，可见在太空中，航天员的行动会受到多么大的限制。而脑机接口技术，成为解决这一问题的可能技术之一。

通过脑机接口技术，航天员直接用思想来输出操作指令，既省去了航天员移动手臂去完成操作花费的大量时间，又减少了体力消耗和精神消耗。航天员只要“趟”在空中，就可以完成一系列运动意图的指令输出，从而完成一系列的飞船隔空控制。2016年，中国“天宫二号”和“神舟十一号”载人航天飞船飞行过程中，两位航天员完成人类历史上首次太空脑机交互。这次测试意义重大，为中国载人航天工程的

新一代医学与人因保障提供了关键的科学依据。虽然现今的脑机接口技术受限于速度、容量和传输精度，无法真正应用到航天员身上，但这些局限会随着该技术的发展而逐渐减小。在未来，脑控技术将会给航天员带来更多的帮助，图4-7展示了在未来航天员将可以在舱内脑控机器人完成舱外的工作的概念图。

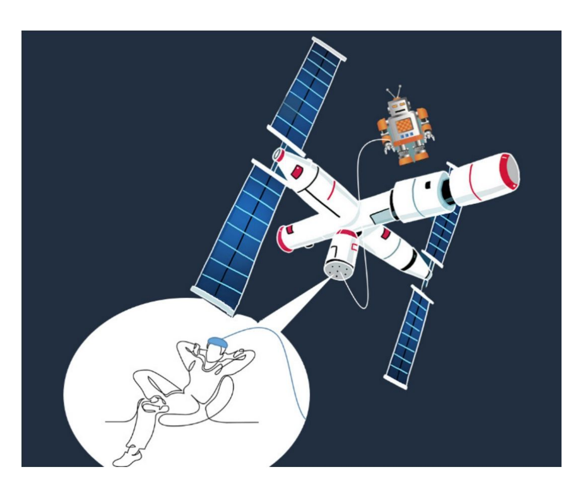


图4-7　航天员未来可以在舱内脑控机器人完成舱外的工作

就在2020年，中国国家自然科学基金委员会批准了“基于双模态脑信息融合的精细关节运动想象解码研究及脑-机接口应用”这一研究项目，由中国航天科工集团二院206所着手研究。该项目的研究成果将填补人体复杂精细运动意图识别的理论和技术空白，为推动脑机结合精细运动意图解码技术在康复工程中的实际应用奠定基础，为实现“所思即所动”的人机绝对运动同步系统探索新思路和新途径。

## 娱乐领域中的脑机接口

### 当人的头上长出了“萌萌的猫耳朵”

在纷繁复杂的社交场合中，人们常常会感慨：长恨人心不如水，等闲平地起波澜。的确，随着年龄的增长，人们会渐渐地开始掩盖自己真实的想法，或是因为羞于表达，或是为了不让对方失望。小明就

时常有这样的烦恼。

小明是一个非常爱说话的同学，经常与朋友坐在一起侃侃而谈，朋友也会在恰当的时候回应他。但有时候，朋友内心所想可能是这样的：你说的这些对我而言没有丝毫的趣味，我回应你也只是因为顾及朋友之间的面子。这种时候，小明常常是一直没有发现而一直讲下去。最后，小明的朋友们因为不愿忍受这种感觉而渐渐地和小明的距离越来越远。小明为此很苦恼，后来小红送给了小明一个带有猫耳的发箍，并告诉他在和朋友聊天的时候可以一起戴上，并在聊天的时候注意观察对方发箍上的两只猫耳朵。后来，小明在和朋友交谈的过程中通过观察猫耳的变化，第一时间看到了朋友们的情绪和态度，及时地转移话题。因此，朋友们深切感受到了小明的体贴，逐渐愿意和他待在一起。而小明，也因为善于观察朋友们的喜怒哀乐，身边的朋友也多了起来。

你是不是觉得这个发箍非常神奇呀？它竟然可以像二次元的小猫咪一样直率、可爱地表达出你的情绪。其实这个神奇的猫耳发箍在现实生活中已经实现了。近年来，日本的神念科技公司就推出了一款头戴式猫耳发箍，叫作“意念猫耳”（图4-8）。外观上看起来就是一个装饰有猫耳朵的发箍，但是实际上，这个小小的发箍里却具有可以读取佩戴者思想和精神状态的“黑科技”。当你对事物充满兴趣并专注于其中的时候，两只猫耳会竖起来；当你心情愉悦的时候，两只猫耳会来回摆动；当你情绪低落陷入悲伤时，两只耳朵也会随着你的情绪耷拉下来；当你身心疲倦提不起精神的时候，猫耳也会跟着你一起“躺平”。

那么这样神奇的猫耳朵是怎么实现的呢？



图4-8　“意念猫耳”可以根据人的情绪不同展现出不同的状态

实际上，“意念猫耳”本质上是一个读取和分析人类脑电波信号的脑机接口系统。大脑进行思考、情绪和各种行为时，数以万计的神经元协同放电产生电信号，这类电信号可以在头皮表面由电极采集到。通过高敏电压表，可以实时地获取头皮电极处的电位信息，这种信息被称为脑电波信号。人在不同的精神状态或是进行不同的心理活动时大脑产生的电信号也是不同的。所以通过分析头皮电极处采集到的生物电信号，就可以判别出人的所想。

“意念猫耳”一共包含3个电极：一个位于额头处（这里是与人情感活动相关的脑区所在位置），另外两个分别位于人的两个耳垂处（作为参考电极）。采集到的脑电波信号被传输到发箍的内置芯片中进行处理。该芯片集成了情感运算库，可以分析出佩戴者的注意力和放松程度，并将其量化，并进行打分。根据数值大小，对佩戴者的精神状态进行定性分析，按照响应的性格特征，对外面的两只猫耳朵发送控制指令。比如当芯片检测到佩戴者的得分为90分，认为他处于注意力非常集中的状态时，会对两只猫耳发送“竖耳朵”的指令，发箍上的猫耳就会通过电机控制竖起来，给人以高度兴奋的感觉。

## 当“脑控”走进元宇宙空间

在信息技术发达的今天，手机、计算机等电子设备让很多人沉浸

其中，网上冲浪，手机购物等，都会给人一种身在家里，心已飞到远方的感觉。《攻壳机动队》《头号玩家》等脍炙人口的科幻作品，给观众们创造了一个个神奇的科幻世界。于是，“元宇宙”这一概念被提出（图4-9）。“元宇宙”是利用科技手段进行链接与创造的，与现实世界映射与交互的虚拟世界，为人们提供了一种新鲜的、低成本的休闲娱乐体验。如何更好地与虚拟世界交互成为当今科技研究的一大热门。为了更进一步提升人们在虚拟世界中的体验效果，目前，一种非常直接的交互方式被提出，即“虚拟现实”。

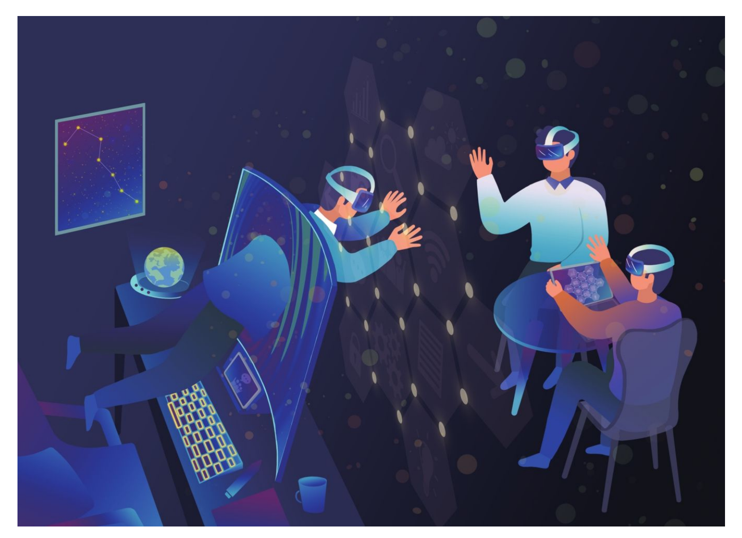


图4-9　元宇宙为人们的业余生活带来了全新的体验

虚拟现实技术是一种利用计算机生成一种可直接对参与者施加视觉、听觉和触觉感受，并允许其交互地观察和操作虚拟世界的技术。例如，当带上虚拟现实眼镜时，视野内将完全变成用计算机设计好的虚拟世界中，给人一种身临其境的感觉。

同样作为一种更为直接的人机交互方式，脑机接口也在被越来越多地应用在虚拟现实技术当中。比如，在虚拟现实康复方面，患者可以通过监测和控制动画运动来重新训练大脑区域。虚拟现实技术也应用于设计和评估基于脑机接口的假肢，帮助完成日常生活需求。此外，虚拟现实可以为适应现实世界场景的程序提供良好的测试场地，

特别是残疾患者可以在过程中学习控制自己的动作或执行特定的任务。

元宇宙在人们的日常生活也有着很大程度的应用。目前，人类就可以利用脑机接口技术操作博物馆向导达到线上参观博物馆的目的。人可以利用事件相关电位信号来控制机器人的导航，用户可以获得一种远程游览的感觉。在新的图形用户界面中，通过聚焦于闪烁的导航箭头来选择命令。

为了简化用户界面，设计者将选择过程分为两部分。每个部分都由不同的事件相关电位诱发。第1部分是从输入阶段开始之前。在这种情况下，用户被要求在两个机器人之间进行选择：机器人1和机器人2决定想要到达的地方，机器人1位于计算机科学系，而机器人2位于植物园。这2个机器人都配备了移动车轮、微控制器、红外传感器、避免碰撞的声呐环和摄像机。即一般来说，第1部分可以被认为是从2个机器人中选择的，机器人1和机器人2位于2个不同的位置。用户可以根据自己的喜好选择机器人来参观，在选择机器人后，使用屏幕给出导航指令，使用停止按钮停止。所有这些都是通过基于事件相关电位的大脑信号来控制的。屏幕会显示机器人摄像机生成的内容。

另外，在休闲娱乐方面，随着脑机接口技术的融入，虚拟现实游戏中的用户体验也会有大幅度改善。其中，智能球游戏旨在降低压力水平，用户通过放松来移动球，从而学会控制他们的压力。驾驶直升机的游戏可以让用户控制飞机飞行到虚拟世界中的任何一点，体验飞行的乐趣。对于角色扮演类的虚拟现实游戏，用户也可以在具有稳态视觉诱发电位的沉浸式三维游戏环境中可以实现对动画角色的控制。也有一些较为休闲类的游戏，用户也可以通过大脑控制实现艺术设计方面的操作，如绘画、涂色等。

传统角色扮演类游戏中，玩家角色的大多数动作都是系统预设的，玩家通过物理按键来对游戏角色的动作进行操纵，实现有限的交互。而在脑机接口技术支持下的游戏可以实现玩家对角色的自由控制。玩家在元宇宙中可以像现实世界一样用自己的意志控制身体每一个部位的活动，实现与游戏世界的自由交互。用过虚拟现实的玩家们应该了解，在进行游戏时会产生眩晕感，这是因为虚拟现实世界中的

虚拟物品缺乏实体导致视觉和触觉产生割裂。而基于脑机接口技术的元宇宙游戏中，由于脑机接口信号的双向传输，玩家会对虚拟世界产生实体感触，你可以感受到晴天时太阳对身体的炙烤，也可以感受到在雨中奔跑时雨点对身体的拍打。

在元宇宙里，玩家的“五感”都可以得到实现。终有一天，基于脑机接口的“元宇宙”，将不再只是一种想象、一种产品、一个空间，而是会成为一种新的“现实世界”。

## 小结

在本章中，我们首先介绍了什么是脑机接口技术与脑机接口技术的发展史，接着介绍了脑机接口技术在医疗领域、军事领域和其他领域的应用。

在医疗方面，脑机接口技术既可以帮助失明患者重新“看见世界”，又可以帮助听力残障患者重新“听见世界”，还可以帮助四肢残障患者重新获得运动能力；在军事方面，脑机接口技术已经成为世界各大军事强国的竞技场，哪个国家能在脑机接口领域取得突破，哪个国家将能在未来战场占据先发优势；在其他领域，脑机接口技术正在成为未来元宇宙空间的入口，在未来，人们将可以通过脑机接口技术进入在元宇宙空间休息、娱乐和生活。

# 5 类脑智能发展的人工智能时代

在漫长的历史岁月中，我们一直认为是心在主宰自己的思想，就像人们常说的“心满意足”“心想事成”“得心应手”……直到人们开始关注大脑，脑科学这一自然科学的“最后疆域”逐渐揭开面纱。随着脑成像技术、大数据、人工智能（artificial intelligence，AI）等领域的快速发展，世界各国对脑科学的研究和探索也愈发激烈。在2021年，科技部网站发布通知称科技创新2030“脑科学与类脑科学研究”重大项目年度申报涉及了59个研究领域和方向，国家拨款经费预计超过31.48亿元，脑科学研究已经成为科技战略重地。

人工智能被认为是21世纪三大尖端技术之一（其他两项为基因工程和纳米科学）。这一浪潮席卷了全球，2018年9月在上海举行的“2018年世界人工智能大会”吸引了来自40个国家和地区的数千名与会者以及数百家大小企业。2022年9月1—3日，以“智联世界、元生无界”作为主题的2022世界人工智能大会在上海世博中心成功举办，此次大会累计举办活动达121场，召集学界、商界以及国际组织等领域500余位重量级大咖，总在线观看人次超6.38亿，盛况空前。可以说，几乎在社会所有领域（金融、制造、教育、通信、医疗、服务等），人工智能都在以前所未有的速度发展着。说到这里，你脑海中有没有浮现出一些你见过的人工智能呢？让我们跟随本书的脚步走进第5章一探究竟吧！

## 跨入人工智能时代

从20世纪被提出到现在，随着相关理论和技术的日渐成熟，“人工智能”已逐步成为一个独立的学科，取得了长足的进步，不仅涉及计算机科学，更涉及心理学、语言学、脑科学等多个学科。“人工智能”的应用体现在日常生活和前沿科技的方方面面，如智能推荐、机器人学、语言和图像处理、博弈、遗传编程等。

## 人工智能促进美好生活

谈起人工智能学科，就不得不提到达特茅斯会议。1956年8月，在美国汉诺斯小镇的达特茅斯学院（图5-1）中，达特茅斯学院的数学系助理教授约翰·麦卡锡（John McCarthy）、人工智能与认知学专家马文·明斯基（Marvin Minsky）、信息论的创始人克劳德·香农（Claude Shannon）、计算机科学家艾伦·纽厄尔（Allen Newell）、诺贝尔经济学奖得主希尔伯特·西蒙（Herbert Simon）等各个领域的佼佼者聚集在这里，探讨一个完全不食人间烟火的主题——用机器来模仿人类学习以及其他方面的智能。虽然很多内容没有达成一致，但这次会议讨论的主题名称被确定下来，即“人工智能”。因此，1956年也被称为“人工智能元年”。此会议上对人工智能的描述为：“如果智能的方方面面都能在多尺度进行精确的描述，而计算机系统能够去模拟的话，那称具有这样能力的系统为人工智能。”



图5-1　达特茅斯学院

今天，“人工智能”不再陌生，日常生活中已随处可见。比如人脸识别、无人驾驶等，用到了计算机视觉相关技术；语音识别、自动翻译用到了自然语言处理相关技术；个性推荐及广告营销（如浏览器推荐新闻、社交网络推荐好友、购物网站推荐服装日用等）等，则涉及数据挖掘等方面的内容；垃圾邮件和垃圾短信的分类与拦截与分类算

法有关。你是否恍然大悟："哦，原来这些都是人工智能！"人工智能发展迅猛，为人类生活带来了更多的便捷。接下来，重点说说日常生活中应用比较广泛的人工智能。

不知道大家平时是否有使用短视频、微博、小红书以及其他购物软件的习惯，有没有发现用过一段时间之后平台推荐的内容越来越符合自己的品味与喜好，有时候甚至还没有等我们主动搜索，平台已经自动将想要的一些好物推送到了首页，这是为什么呢（图5-2）？难道我们被平台"监视"了吗？其实这主要是推荐系统发挥了重要作用。推荐系统是一种特别的信息过滤系统，可以根据现有的用户信息和数据（包括搜索、评分、点击率、停留时长、互动情况等）进行个性化推荐内容。可以这么想，推荐系统根据不同人的使用情况为每个人制订一张独一无二的"自画像"，这张图像包含体现着我们的喜好、习惯等信息。也可以这样理解，每个用户都有一个专属的"浏览管家"，根据不同的客户实现不同的推送结果。亚马逊就是通过这类推荐引擎达到营业收入提高35%的业绩。与此类似的案例还有个性化穿搭、个性化健身、个性化医疗以及特定风格音乐、书籍、新闻推送等。

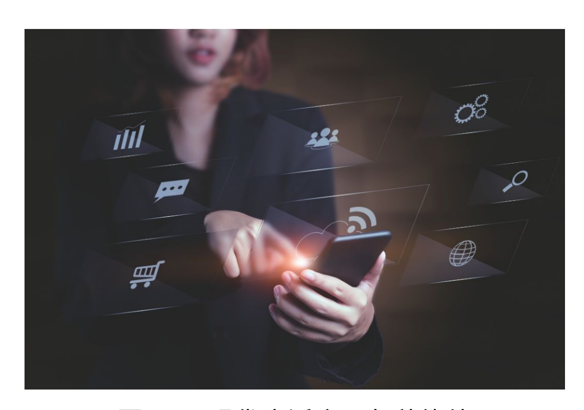


图5-2　现代生活应用各种软件

从最开始的只能翻译文字到现在可以实现语音（图5-3）、图片、文件直接读取翻译，从最开始需要设置语言类别到后来的能够自动识别语言体系，翻译软件的不断更新进步为很多学生和工作人员提供了更大的便捷和更好的使用感。翻译软件的背后，是人工智能中的自然语言处理技术，简单来说，就是将可识别的输入文本通过自然语

言处理算法输出为特定结构化的数据，其他的应用还包括声音转换、评论或者话题的情感分析，等等。



图5-3　语音翻译的使用场景

随着社会普遍安全意识的增强，很多小区和公司都在入口设置了面部识别（图5-4），有效提高安保效率；大家排队做核酸的时候刷身份证进行快速登记；无人机在例行检查时可以实现区域搜索、障碍判断等，从而帮助绘制地图……这一系列的使用场景都是在计算机视觉的支持下完成。此外，该技术在唇语识别、视频文字提取、图像分类、机器人控制等方面也有着广泛的应用。

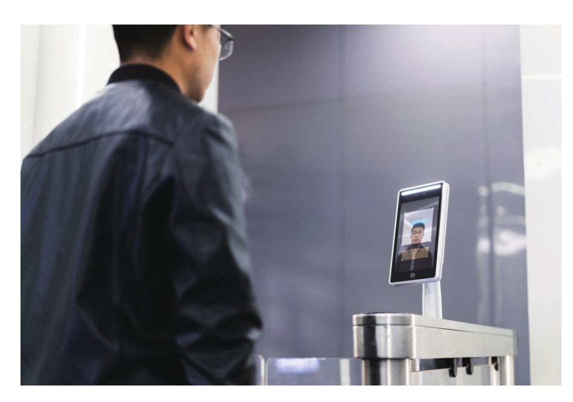


图5-4　面部识别的使用场景

在2022北京冬奥会中，人工智能的使用无疑给冬奥会加持了浓重的科技色彩，成为北京冬奥会的亮点之一，有负责热红外测温、口罩

佩戴检测、公共空间巡控、手部消毒等工作的巡检机器人，也有负责餐饮工作的后勤机器人，还有穿梭于场馆闭环与非闭环区的无人配送车等。另外，人工智能还可以助力运动员动作技术的分析，帮助运动员科学提升训练水平。科技冬奥重点专项“冬季项目运动员专项能力特征和科学选材关键技术研究”课题负责人、北京体育大学运动与健康研究院院长刘卉教授团队利用基于深度学习原理的人工智能技术，解决了“跟得住”“识别准”“精度高”3个主要问题，可以进行大范围的高空动作数据采集，实现对视频中人体关节点的自动识别，进而建立起适用于竞技体育和一般生物力学研究的计算机系统——无反光点人体运动自动捕捉人工智能系统。该系统已应用在了速度滑冰、花样滑冰、跳台滑雪、越野滑雪、钢架雪车等项目的训练工作中。其中，在速度滑冰与越野滑雪训练中，该系统已经获取超过8000人次的动作数据。

现在市面上非常火热的家居产品，有很多都主打“智能家居”的主题，如智能窗帘、智能灯光控制、智能扫地机器人（图5-5）以及智能安防。前两款可以根据用户是否居家和具体要求切换不同的情景模式，比如不在家的时候可以换为离家模式，实现家居自动关闭的功能。智能安防是普遍反响比较好的一款系统，尤其家里有老人或者孩子，用户可以通过智能监控实时以及延时查看情况，也能够进行对话，十分便捷，让用户无论何时何地都可以对家庭情况了如指掌。



图5-5　吸尘器机器人

随着新冠肺炎疫情在全世界范围的扩散，人工智能在防疫方面也

展示出了不可忽视的重要使用价值。医院作为可能引起疾病传播的重点公共场所之一，防疫和人员管控至关重要。为筑牢疫情防控屏障，巩固来之不易的防疫成果，商汤科技与上海中医药大学附属曙光医院合作开发了智能流调系统。该系统将商汤科技的“企业方舟开放平台”和曙光医院研究的流行病学调查参数指标系统进行结合，通过高精度的人工智能视觉感知，自动流调相关人员的时空行程与密切接触情况，并且还可以显示流调对象和其周围人员是否规范佩戴口罩。这一技术解决了医院传统流调工作中耗时费力与效率瓶颈的问题。人工智能在追踪接触者、快速筛查和预测疾病发展过程等方面发挥了积极且十分重要的作用。

可以看到，人工智能已经渗入我们生活的众多领域，在很多方面提高了便利性和效率，促进了广大人民的美好生活，提升了幸福指数。人工智能带来的巨大变革不仅体现在日常生活中，在科研领域也有许多杰出的成绩。

## 人工智能助力科技腾飞

我们知道，人工智能领域的神经网络很大程度上是模仿人类神经元建立起来的，甚至“神经网络”这一名词都是从生物学界引用而来的。那么，在神经网络的雏形之上，科学家们是怎样一步一步地搭建起今天所见的强大而高效的人工智能的呢？科学的发展往往是螺旋上升的，有繁荣和突破，也有冷遇和停滞不前，人工智能的发展历程也是如此。

1980年，日本科学家福岛邦彦创造性地从人类视觉系统引入了许多新的思想到人工神经网络，搭建了一个全新的神经网络模型，被很多人认为是如今广泛应用的卷积神经网络的雏形。有趣的是，福岛邦彦的初衷是构建一个像人脑一样，能够识别看到的物体的网络，来帮助我们更好地理解大脑的运作，却无意间为现代人工智能的发展奠定了基础。

接下来的10年间，关于卷积神经网络的研究始终停滞不前，直到1990年，科学家杨立昆（Yann LeCun）在福岛邦彦的基础上引入了

新的反向传播算法，并且简化了卷积运算的过程，使卷积神经网络初步具备了大规模应用的基础。但这位科学家并没有止步不前，1998年他再次发表了一篇长达46页的论文，提出了一个新的网络模型，并且将自己的方法与当时全部的主流机器学习方法做对比，取得了压倒性的胜利。事实上，这个被命名为“LeNet-5”的网络在基础架构上已经无限接近今天的卷积神经网络了。人们本以为这是人工智能崛起的冲锋号，但由于当时计算机的计算能力较弱，无法训练大规模的神经网络，人工智能的发展在世纪之交又一次陷入了迷失。值得一提的是，杨立昆教授于2019年获得了计算机领域的最高奖项——图灵奖。

直到2012年，在充分进步的硬件计算能力支持下，多伦多大学的亚历克斯·克里泽夫斯基（Alex Krizhevsky）等搭建了比以往的神经网络都要更深的网络，在图像分类的任务中取得了压倒性的优势，令当时所有的人工智能方法望尘莫及。时至今日，这个模型的提出已经被公认为人工智能发展的里程碑。而正因为克里泽夫斯基等搭建的网络，就其本质而言，是以往的神经网络的深层版本，当时的人们发现了深层网络的巨大潜力，并引发了关于“越深越好”的思考，这也成为深度学习蓬勃发展的开端。

发人深省的是，克里泽夫斯基开始钻研人工智能时，已经即将从多伦多大学毕业了。面对毕业前的最后一份工作，他并没有敷衍了事，而是以过人的毅力和创造性的思维，展现了人工智能的广阔前景，为人类的科技进步做出了卓越的贡献。

2014年，依托于博弈论思想，兰·古德费洛（lan Goodfellow）搭建了“生成对抗网络”。今天我们所见到的人工智能，已经能够根据描述生成逼真的人脸图像，对真实的人脸图像进行风格上的转变。甚至，人们只需要指定一种风格，比如“学生”，人工智能就可以生成成千上万的可以以假乱真的“学生”的高清图像，这些强大的功能都是在“生成对抗网络”的基础上实现的。

时间来到2015年，科学家们已经在“越深越好”的道路上遇到了重重阻碍，他们发现，人工智能的规模越大，神经网络的层数越深，训练就越艰难，而取得的效果也很难令人满意。更令人迷茫的是，随着网络层数的加深，很多人工智能模型的能力居然发生了退化。深度学

习该如何发展？深度学习是否还有未来？这两个问题在当时引发了大规模的讨论。在这个决定人工智能何去何从的十字路口，华人科学家何恺明带着他的“深度残差网络”横空出世，一锤定音地开启了人工智能的黄金时代。“深度残差网络”的提出，使得深达几十层甚至上百层的神经网络依然可以被训练和应用。这是第一个在图片分类任务上超越人类的人工智能模型，也是第一个在工业界繁荣发展，被大规模应用于各种科技产品中，从方方面面改变人类生活的人工智能模型。时至今日，“深度残差网络”依然运行在世界各地的人工智能研究机构的计算机中，运行在全球各大科技公司的产品中，它的各种变体已经成为人工智能领域的通用框架。同时“深度残差网络”也是人工智能学术研究中的一个标杆，科学家们每提出一种新的人工智能模型，都要首先证明：我们的方法并不弱于数年前的深度残差网络。

2016年，人工智能终于以一种别样的方式出现在大众的视野中。在“深蓝”打败国际象棋大师加里·卡斯帕罗夫（Garry Kasparov）后，人工智能向人类棋类运动的最后一块领地：围棋发起了冲锋。在2016年之前，大众普遍认为，由于围棋运动更依赖于人类玩家的直觉，人工智能挑战顶尖的人类选手还需要很多年。但得益于深度学习和树形搜索策略，阿尔法围棋（AlphaGo）首次实现了人工智能对人类选手的胜利（图5-6），并使得“人工智能”这个名词与科学研究脱钩，以一个崭新的科技产品的形象为人所熟知。

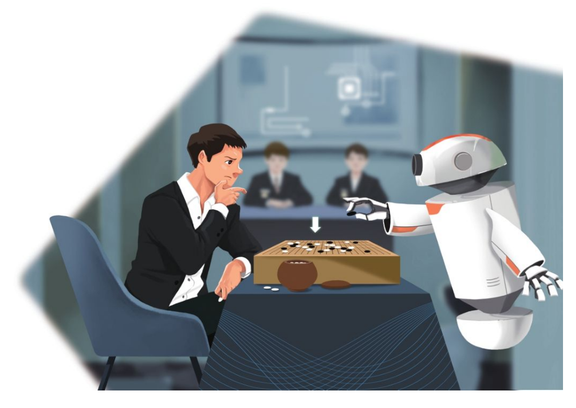


图5-6　人工智能与人类围棋博弈

2017年，科学家们将目光转向了人类理解机制中的一个重要部分：自注意力。他们观察到，人类在理解文字或图像时，往往能捕捉到文字或图像内部各个成分间的联系，并将关注点放在与其他成分联系最紧密的少数几个组成部分上。受此启发，科学家们将这种注意力机制引入人工智能领域，使得人工智能在自然语言和图像理解上取得了重大突破。今天，我们可以与人工智能对话，可以让人工智能理解我们的语音指令，可以让人工智能根据上下文补充文章中的缺失部分，甚至可以让人工智能进行文学创作，这些科技进步都得益于自注意力机制的提出与发展。

过去10年，是人工智能发展历史上一个令人难以置信的高速发展和多样创新的时期，许许多多的科研成果不断颠覆传统认知，许许多多曾被认为是天方夜谭的科技产品走进日常生活。随着人工智能的不断发展，人们对于“智能”的理解也越来越深刻。我们不禁期待，未来人工智能又会取得怎样的累累硕果呢？

## 类脑智能与未来

正如第1章中所描述的那样，“中国脑计划”的其中一个应用层面是脑机智能技术，面向类脑智能产业，主要包括两点：脑机接口和类脑研究。脑机接口主要研究大脑和机器之间的联系，在本书的第4章已经进行了详细介绍，这里就不再赘述。类脑研究是未来智能的基础，是下一代人工智能需要的理论研究。可以说，类脑智能是迈向未来智能的重要关卡。

### 人工智能与人脑的“执子之手，与子偕老”

事实上，人工智能领域的突破离不开脑科学的启发。许多先驱的人工智能科学家也是脑科学家，如艾伦·图灵（Alan Turing）、约翰·麦卡锡（John McCarthy）、马文·明斯基（Marvin Minsky）。大脑之间的神经连接启发计算机科学家开发了人工神经网络；大脑的卷积性质和多层结构又启发了研究人员开发卷积神经网络和深度学习；受到自注意力机制的启发，人工智能在自然语言和图像理解上取得了重

大突破……同样，人工智能的发展也使得人类更加关注大脑，促进了脑科学的进一步发展。



脉冲神经网络是源于生物启发的新一代人工神经网络。尽管长久以来深度神经网络凭借计算机的强大算力在很多领域都有所突破，但并不高效。于是人们关注到生物神经的编码方式是离散的脉冲形式，不同脉冲出现的时间序列也是编码信息的重要组成部分。因为大脑动态神经网络中的神经元并不是在每一次信号迭代传播中都被激活，而要在它的膜电位达到阈值才被激活，从而产生脉冲进而再恢复静息膜电位。因此，如果达不到阈值，那么神经元就不会有脉冲发生，膜电位也保持不变。由此人们受到启发，发展出了第三代神经网络——脉冲神经网络。我们可以看出，脉冲式编码更加符合神经元真实的工作状态，这也使得编码更加轻松与自由。

曼彻斯特大学研发的SpiNNaker，号称“世界最大的‘大脑’”，拥有100万个处理器核心和1200块互连电路板，希望通过模拟人类大脑的行为帮助我们更好地理解大脑的运行机理以及与大脑相关的疾病，例如帕金森病、阿尔兹海默病等。



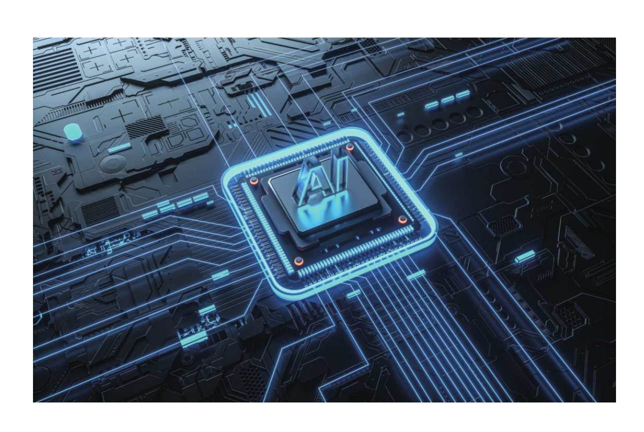
由上可以看出，人工智能与人脑存在着密不可分的关系，相互促进。但人工智能的发展，也引起了社会对人机关系和相处模式的思考。无论是上面提到的IBM开发的象棋计算机“深蓝”战胜棋王加里·卡斯帕罗夫，还是谷歌阿尔法围棋战胜围棋九段柯洁，又或者是Deep Stack战胜德州扑克人类职业玩家……伴随着一次次人工智能与人脑的博弈，人们既对它的前景充满期待，与此同时各种舆论也甚嚣尘上：人工智能会取代人脑吗？人工智能能否与人类和平共存……

不知道大家是否看过电影《人工智能》，影片中母亲莫尼卡（Monica）决定收养一个机器人小孩大卫（David），她通过念词程序启动了大卫对自己的爱，从此母亲莫尼卡成为大卫生存的唯一理由，是大卫生命里真正的一束光。而机器人大卫也为莫尼卡带来了前所未有的快乐，治愈了她忧虑的心灵。但最终随着莫尼卡不再信任大卫，为了保护她真正儿子的安全，这个机器人“儿子”被遗弃。其实这部影片主要是为了揭示一个人文核心问题——到底什么才是人类的本质？如果机器人能够与人类无异，具备思考、学习、爱与仇恨等各种高级能力，人类是否会感到威胁？

虽然AI对人脑的挑战不会停下，但以现在的发展情况来说，断言人工智能会取代甚至毁灭人脑还言之过早。计算机的出现使得人类开始模拟大脑智慧，人工智能可以说发展到了遍地开花的程度，众多产品已经闯入了千家万户。但人脑的“智能”和这些人工智能是同样的吗？

就目前而言，人工智能还没有达到真正的“智能”，还无法实现像

人脑一样思考、运作。尽管人工智能被用于人脸识别、文字识别或下棋等某些目标和规则明确的任务，其在计算速度和准确性方面超越了人类。但这更多的是从大量数据当中寻找规律，机械地对数据拟合与有限的泛化，称其为“数据智能”似乎更加恰如其分。




而人类的大脑是一个复杂的系统，可以根据少量数据得出复杂结论，能够同时并行处理很多事情，面对新事物能够产出新的知识，能够根据自己的喜好做出选择，具有同理心、情绪、好奇心、创造力、终生学习的能力……人类的大脑是很灵活的，尤其是可以做到举一反三，比如看到外面下着大雨，你不会毫无准备立刻出门，要么拿伞，要么等雨停；天冷了多加衣服；热水要晾一会儿再喝……

这样一个个十分日常又简单的行为背后蕴含着复杂的大脑运行机制，但这些对于目前的人工智能来说是难以表达和实现的。换句话说，人类智能的本质在于不断适应环境并能在现实世界中行动与生存的能力。因此，人类智能和我们通常所说的人工智能属于两类不相同的智能进化形态。可以说，人工智能还处在非常初级的阶段，达到与人类智能同等水平的机器人仍然处在科幻小说的世界里。

## 人工智能的下一个“春天”——类脑智能

可以看出若是停留在行为尺度的模拟，这与构造真正意义的智能是不同的。“机器能够思考吗？”艾伦·图灵在20世纪50年代就有此类的思考。之后，美国计算机学会的创会主席埃德蒙·伯克利（Edmund

Berkeley）在他的著作中也曾提到“CAN MACHINES THINK? WHAT IS A MECHANICAL BRAIN?”可以看出，机器能够具有思维的能力应该是当时探索计算机作为机械大脑的一批优秀科研人员的美好愿景。

过去脑科学中一些复杂的结果，并没有被应用到人工智能中，甚至一些十分简单的原理都还没能在人工智能中实现。就像人类大脑的联接是动态变化的，有的可以生成，有的可以消减。然而，在人工神经网络中所有的联接都是固定的。这种类似的简单脑科学原理应用在未来智能都将产生十分大的影响。

中科院自动化所所长徐波研究员在CCTV-2《对话》节目中访谈时表示，下一代人工智能应该具备三个特点，低功耗、具有自主学习能力、在价值观上实现人机协同。第一，在低功耗方面，虽然现有的人工智能模型结构上部分借鉴了大脑的神经形态，但是它的学习方法上还主要是基于一种叫“误差反传”的数学最优化方法，这就使得能量消耗十分庞大。比如说最近发展出来的大模型技术，训练一个这样的模型碳排放相当于一辆小汽车从地球到月亮的一个来回，而我们人类大脑的能耗仅仅在20瓦左右，二者之间能耗差距巨大。第二，在自主学习方面，上面也提到了现有的人工智能依赖于大量的数据进行封闭式学习，实现自主学习、举一反三等能力十分困难。第三，如果能够在人工智能中实现人类价值观的认同，将会给产业变革带来重要影响。要进一步在人工智能领域实现“里程碑式”的进展，“类脑智能”接过接力棒，成为人工智能研究的“新宠”。

类脑智能是以计算建模为手段，受脑神经机制和认知行为机制启发，并通过软硬件协同实现的机器智能。类脑智能系统在信息处理机制上类脑，认知行为和智能水平上类人，其目标是使机器以类脑的方式实现各种人类具有的认知能力及其协同机制，最终达到或超越人类智能水平。类脑智能是一个交叉学科，需要脑科学、认知科学、算法、硬件、心理学等多种学科的深度融合，它有望弥补传统人工智能的不足，带领人工智能走向下一个春天，迎来技术奇点。

在类脑研发领域，中国一些技术成果走在了前列。例如类脑芯片，清华大学研制的“天机芯”问世，实现了中国在人工智能和芯片两大领域在《自然》上杂志发表论文零的突破。什么是类脑芯片？目前

人工智能中神经网络模型的最重要问题就是计算量大导致的算力需求快速增长和算力提升放缓的尖锐矛盾。面对这样的现实环境，我们以期通过类脑芯片解决。类脑芯片是人工智能芯片中的一种架构，模拟人脑进行设计，一旦信号开始在它的“血管”里流淌，就能像生物的大脑一样进行思维，并做出反应，在功耗和学习能力上具有更大优势。例如，阿尔法围棋与人类进行围棋大战时，需要耗费将近1000度电，但是采用类脑芯片后，会大大降低能耗，估计仅用原来能耗的1/300就可以完成同样的工作。而与此同时，运算速度却能达到原来的上百万倍甚至上亿倍。



中国清华大学类脑计算研究中心施路平团队研发的类脑芯片第三代“天机芯”，做到了脉冲神经网络和人工神经网络的兼容并蓄，是全球首款异构融合类脑芯片。第三代“天机芯”，包含约40 000神经元和1000万突触，搭载“天机芯”的无人驾驶自行车，实现了语音理解控制、自动避障、自主决策等功能，大家可以在网络上搜索“天机芯”应用于无人自行车行驶的展示视频。我们可以简单理解这两类神经网络的分工：脉冲神经网络负责语音识别、以及对不同神经网络的整合与决策等功能，而人工神经网络负责实现图像识别和物体检测、控制平衡和方向、人类目标跟踪等功能。美国《纽约时报》对此成果评论：这可能是“最接近自主思考的无人驾驶自行车”。“天机芯”是中国在智能芯片研发史上的重大成果，具有里程碑式的意义。

除此之外，2015年浙江大学也推出了其自主研发的“达尔文一

代”芯片，并于2019年推出“达尔文二代”芯片；2020年，浙江大学集成792颗达尔文二代芯片组成了一台类脑计算机。在2018年的“神经启发计算元素研讨会”上，海德堡大学推出Brain Scales芯片，工作速度比普通芯片快1000～10 000倍，第二代Brain Scales芯片具有片上学习功能。此外，英特尔Loihi芯片、高通Zeroth芯片、西井科技Deep-South芯片、AI-CTX芯片也都在类脑芯片上努力。不过这些产品距离大规模商业化的程度仍然很远。

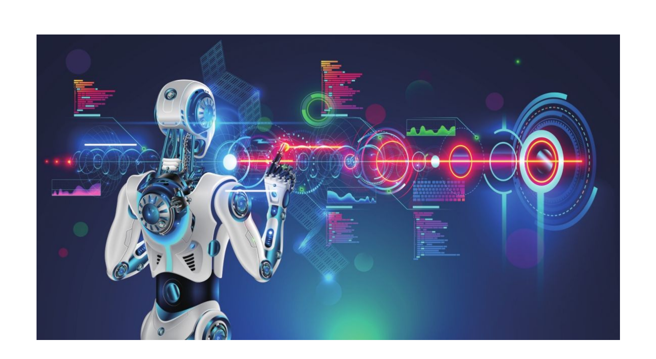
我们需要知道尽管在类脑芯片领域已经取得了很多令人瞩目的成果，但这些与人脑的工作模式还存在很大差距。除了类脑芯片，类脑智能未来的发展重点方向还包括脑机接口、类脑智能机器人、机器学习、认知计算、混合现实等。

目前类脑智能整体处于实验研究阶段，中国各科研机构与企业也大多还是处在起步、争相发力的阶段，真正实现相关技术商业化应用还有很长的路要走。要实现真正的“智能”，还需要更多理论的研究与技术的进步。

## 小结

在本章中，我们首先介绍了人工智能在生活和科研领域的发展与应用，接着探讨人工智能与人脑之间的关系，介绍了类脑智能的现状，最后畅想类脑智能的未来。

类脑智能是一个交叉学科，需要脑科学、认知科学、算法、硬件、心理学等多种学科的深度融合，未来类脑智能的发展与进步需要理论研究的新发现、软硬件层面的新突破以及产品层面的最终落地转化。



# 结语

现在，亲爱的读者，大脑探索之旅程即将结束。

脑科学被称为科研领域“皇冠上的明珠”，是研究大脑结构和功能的科学，是理解自然和人类本身的“最终疆域”，是生命科学最难以攻克的领域之一。

通过阅读本书，我们知道了大脑是人体最复杂的器官，包含着上百亿个甚至千亿个神经元，彼此之间通过突触连接等方式构成了一个庞大而又复杂的神经系统，完成思考、记忆、注意、认知控制等成千上万的事情，神经系统损伤则可能带来不同类型的困扰。“我们是如何看到的”“我们是如何听到的”“我们是怎么记住一件事情的”“意识存在于何处”“喜、怒、哀、惧、爱、恶，人的七情六欲又由大脑的哪些区域控制”，聪明的你，在书中找到答案了吗？

面对脑科学这一仍未被完全开垦的领域，大国正纷纷起跑，先后启动针对大脑的研究项目。中国也于2021年正式启动“脑科学与类脑科学研究”，即“中国脑计划”。时至今日，理解脑的工作机制，对于重大脑疾病的早期预防、诊断和治疗，人脑功能的开发和模拟，创造以数值计算为基础的虚拟超级大脑，以及抢占国际竞争的技术制高点具有重要意义。正如文中描述的，人工眼球、人工耳蜗帮助残障人群正常生活；意念控制从科幻正在逐步走入现实；类脑芯片的研究与开发期望达到模拟人脑处理信息的目的，脑科学研究对计算机、人工智能等领域的诞生和发展产生的深远影响正在不断打破人们的思维边界。

在本次旅程结束之前，作者有几句话想说。从“心智源于心脏”到“思维、情感、智慧皆来自大脑”，探索脑科学奥秘的道路并不是一帆风顺，充满了曲折和坎坷。本书仅对目前的研究成果进行概述，脑科学研究仍处于不断探索的阶段，不同观点错综复杂，一些研究领域也会出现争论和分歧，其实，这是正常现象。毕竟认知角度是多元的，不同观点是可以碰撞的，正所谓真理越辩越明。

青少年是祖国的未来，希望这本书可以让广大青少年更加客观地

了解大脑、认识大脑、理解大脑，新一代的研究力量可能就来自各位热爱脑科学的同学们。

# 参考文献


[1]贝尔，科勒斯，帕罗蒂斯．神经科学——探索脑：第2版[M]．王建军，主译．北京：高等教育出版社，2004．

[2]何静．人类学习与深度学习：当人脑遇上人工智能[J]．西南民族大学学报(人文社科版)，2017，38(12):84-88．

[3]贺文韬．脑机接口技术综述[J]．数字通信世界，2018(1): 73-78．

[4]胡剑锋．未来不是梦——脑机接口综述[J]．江西科技学院学报，2006(2): 81-88．

[5]李伟．认知建模和脑控机器人技术[M]．北京：科学出版社，2021．

[6]林涵，石海明，曾华锋．从DARPA资助BCI技术研发看未来军事变革[J]．国防科技，2011，32(5): 52-59．

[7]加来道雄．心灵的未来：理解、增强和控制心灵的科学探寻[M]．伍义生，付满，译．重庆：重庆出版社，2015．

[8]加扎尼加，伊夫里，曼根．认知神经科学：关于心智的生物学[M]．周晓林，高定国，译．北京：中国轻工业出版社，2011．

[9]孟海华．类脑智能的发展趋势与重点方向[J]．张江科技评论，2021(2): 67-69．

[10]地球上的阿葛．升级吧，大脑！给上进青年的用脑指南[EB/OL]．[2022-10-17]．https://www.zhihu.com/remix/albums/1058016904253419520．

[11]Jenny蔡健玲．斯坦福泰斗带你入门心理学[EB/OL]．[2022-10-20]．https://www.zhihu.com/remix/albums/931603443288780800．

[12]肖琳芬．蒲慕明院士：脑科学与类脑智能[J]．高科技与产业化，

2021，27(10):20-23．

[13]杨廙，李东，崔倩，等．触觉的情绪功能及其神经生理机制[J]．心理科学进展，2022，30(2): 324-332．

[14]叶浩生．心理学通史[M]．北京:北京师范大学出版社，2006．

[15] PEI J, DENG L, SONG S, et al. Towards artificial general intelligence with hybrid Tianjic chip architecture[J]. Nature, 2019, 572(7767): 106-111.

[16]于淑月，李想，于功敬，等．脑机接口技术的发展与展望[J]．计算机测量与控制，2019，27(10): 5-12．

[17]王志良．脑与认知科学概论[M]．北京：北京邮电大学出版社，2011．

[18]曾毅，刘成林，谭铁牛．类脑智能研究的回顾与展望[J]．计算机学报，2016，39(1): 212-222．

[19]赵倩，谭浩然，王西岳，等．脑电采集电极研究进展[J]．科学技术与工程，2021，21(15):6097-6104．

[20]张发华，舒琳，邢晓芬．头皮脑电采集技术研究[J]．电子技术应用，2017，43(12):3-8．

[21] BEAR M F, CONNORS B W, PARADISO M A. Neuroscience: Exploring the brain[M]. 4th ed. Philadelphia: Lippincott Williams and Wilkins, 2015.

[22] GU L, PODDAR S, LIN Y, et al. A biomimetic eye with a hemispherical perovskite nanowire array retina[J]. Nature, 2020, 581(7808): 278-282.

[23] MORRISON I, LÖKEN L S, MINDE J, et al. Reduced C-afferent fibre density affects perceived pleasantness and empathy for touch[J]. Brain, 2011(134): 1116-1126.

[24] JENNY, LIU, BETSY, et al. Google Effects on Memory: Cognitive Consequences of Having Information at Our

Fingertips[J]. Science, 2011(333): 776-778.

[25] NEWSOME W T, PARE E B. A selective impairment of motion perception following lesions of the middle temporal visual area (MT)[J]. Journal of Neuroscience, 1988, 8(6): 2201-2211.

[26]NIMET (U)NAY GÜNDO(G)AN, AYSE CANAN YAZICI, AYTEN SIMSEK . 优势眼测量法的研究[J]. 国际眼科杂志, 2008(10): 1980-1986.

[27] SOMERS B, LONG C J, FRANCART T. EEG-based diagnostics of the auditory system using cochlear implant electrodes as sensors[J]. Scientific Reports, 2021,11(1).

[28] POO M M. Towards brain-inspired artificial intelligence[J]. National Science Review, 2018, 5(6): 785.

[29] WICKENS A P. A History of the brain: from stone age surgery to modern neuroscience[M]. Hove, East Sussex: Psychology Press, 2015.


基础前沿科学史
丛书
给青少年讲
量子科学
高鹏 著
清華大學出版社

基础前沿科学史丛书

# 给青少年讲量子科学

高鹏　著

清華大学出版社
北　京

图书在版编目（**CIP**）数据

给青少年讲量子科学 / 高鹏著．—北京：清华大学出版社，2022.10

（基础前沿科学史丛书）

ISBN 978-7-302-61978-9

Ⅰ．①给… Ⅱ．①高… Ⅲ．①量子论 - 青少年读物 Ⅳ．①O413-49

中国版本图书馆CIP数据核字（2022）第181610号

责任编辑：宋成斌

封面设计：意匠文化·丁奔亮

责任校对：王淑云

责任印制：曹婉颖

出版发行：清华大学出版社

网　　址：http://www.tup.com.cn, http://www.wqbook.com

地　　址：北京清华大学学研大厦A座

邮　　编：100084

社总机：010-83470000

邮　　购：010-62786544

投稿与读者服务：010-62776969，c-service@tup.tsinghua.edu.cn

质量反馈：010-62772015，zhiliang@tup.tsinghua.edu.cn

印装者：三河市龙大印装有限公司

经　　销：全国新华书店

开　　本：165mm×235mm

印　　张：15.75

字　　数：170千字

版　　次：2022年11月第1版

印　　次：2022年11月第1次印刷

定　　价：69.00元

---

产品编号：097714-01

# 丛书序　给面向青少年的科普出版点一把新火

2022年是《中华人民共和国科普法》通过的第20年，在这样一个对科普工作意义不凡的年份，由北京市科学技术委员会（以下简称市科委）发起，清华大学出版社组织的“基础前沿科学史丛书”正式出版了。这套书给面向青少年的科普出版点了一把新火。

2022年9月4日，中共中央办公厅、国务院办公厅印发《关于新时代进一步加强科学技术普及工作的意见》，进一步强调“科学技术普及是国家和社会普及科学技术知识、弘扬科学精神、传播科学思想、倡导科学方法的活动，是实现创新发展的重要基础性工作”。科学技术普及是科技知识、科学精神、科学思想、科学方法的薪火相传——是“薪火”，也是“新火”。

市科委搭台，出版社唱戏，这套书给面向青少年的科普图书出版模式点了一把新火。市科委于2021年11月发布了“创作出版‘基础前沿科学史’系列精品科普图书”的招标公告，明确要求中标方在一年的时间内，以物质科学、生命科学、宇宙科学、脑科学、量子科学为主题，组织“基础前沿科学史”系列精品科普图书（共5册）出版工作；同步设计制作科普电子书；通过网络媒体对图书进行宣传推广等服务内容。这些服务内容以融合出版为基础，以社会效益为初心。服务内容的短短几句话，每一句背后都是特别繁复的工作内容。想在一年的时间内，尤其是在2022年新冠肺炎疫情期间，完成这些工作的难度可想而知，然而秉承“自强不息，厚德载物”的清华大学出版社的出版团队做到了。

中国科学家，讲好中国故事，这套书给面向青少年的科普图书选题内容点了一把新火。中国特色社会主义进入新时代，新一轮科技革命和产业变革正在深入发展，基础前沿科学改变着人们的生产生活方式及思维模式。《中华人民共和国国民经济和社会发展第十四个五年规划和2035年远景目标纲要》提出：在事关国家安全和发展全局的基

础核心领域，制定实施战略性科学计划和科学工程。物质科学、生命科学、宇宙科学、脑科学、量子科学等领域，迫切需要更多人才参与研究，而前沿科学人才的建设培养，要从青少年抓起。这5本书的作者都是中国本土从事相关专业领域工作的科学家，这5本书都是他们依托自己工作进行的原创性工作。虽然内容必然涉及科学史的内容，但中国科学家尤其是近些年的贡献也得到了充分展示。

初心教育，润物无声，这套书给面向青少年的科普图书科普创作点了一把新火。习近平总书记提出：科技创新、科学普及是实现创新发展的两翼，要把科学普及放在与科技创新同等重要的位置。因此，针对前沿科技领域知识的科普成为重点。如何创作广受青少年欢迎的优秀科普图书，充分发挥科普图书的媒介作用，帮助青少年树立投身前沿科学领域的梦想，是当前科普出版工作的重点之一，这对具体的科普创作方法提出了要求。这套书，看得出来在创作之初即统一了整体创作思路，在作者进行具体创作时又保持了自己的语言习惯和科普风格。这套书充分体现了，面向青少年的科普图书创作，应该循序渐进，张弛有度，绘声绘色，娓娓道来，以科学家的故事吸引他们，温故科学家的研究之路，知新科学家的科研理念，以科学精神润物细无声。

靡不有初，鲜克有终。2022年10月16日，习近平总书记在中国共产党第二十次全国代表大会报告中强调“教育、科技、人才是全面建设社会主义现代化国家的基础性、战略性支撑”。且将新火试新茶，诗酒趁年华。希望清华大学出版社的这套“基础前沿科学史丛书”为广大青少年推开科学技术事业的一扇门，帮助他们系好投身科学技术事业的第一粒扣子，在全面建设社会主义现代化强国的新征程上行稳致远。

中国工程院院士
清华大学教授



# 前言

近几年，随着量子科技的发展，“量子”一词越来越频繁地出现在大众的视野中。许多人对量子力学充满好奇，但又对量子力学似懂非懂，感觉它非常神秘，网络上甚至出现了“遇事不决，量子力学”的戏谑之语。在大部分人心目中，量子力学是跟高深莫测联系在一起的。新闻报道中“量子计算机”超凡的计算能力让人们对量子科学又敬又畏，而今年刚刚获得诺贝尔物理学奖的科学家所做的“量子纠缠”又让人们感觉云里雾里、十分新奇，甚至有人将其与心灵感应联系在一起，但实际上，今年获得的诺奖的工作早已完成几十年了，由此可见，公众对量子力学的了解还是十分缺乏的。

人类文明发展至今，经历了石器时代、青铜时代、铁器时代、蒸汽时代、电气时代和信息时代。在石器、青铜和铁器时代，中华文明取得了辉煌的成就，一直是世界文明的引领者。但是，从蒸汽时代开始，我们逐渐落后，错过了蒸汽时代和电气时代的发展大潮。好在经过百年奋斗，急起直追，现在的信息时代我们虽然有所落后，但已经逐渐赶了上来。那么，未来的下一个文明时代会是什么时代呢？从目前的科技发展情况来看，很可能是量子信息时代。可喜的是，近年来我国的量子信息科技取得了一系列令世界瞩目的成果，处于领跑的位置，也许在未来的量子信息时代，我们将重新成为世界文明的引领者，这将是实现中华民族伟大复兴的重要一环。各位同学，你们很荣幸，因为你们很可能都将是其中的参与者。

量子力学的发展已经有了一百多年的历史，不管它提出来的时候多么难以理解，但是经过这么些年的发展，它理应成为公众常识的一部分，就像牛顿力学已经成为公众常识一样。而事实上，量子力学的普及程度远远不及牛顿力学，其原因，当然与量子力学与人们的日常生活很难发生交集有关，但是，随着量子信息时代的到来，“量子”将越来越多的出现在人们的生活中，因此，在公众中普及量子科学就成为一件很有必要的事情。作为一个科学与教育工作者和一个为量子着迷二十年的量子爱好者，我很希望尽自己的微薄之力，能为量子科学

的普及做一点小小的贡献，让公众尤其是青少年能尽可能地在不涉及深奥的数学的情况下，读懂量子科学并理解其中的奥妙。

实际上，我并不是量子力学的研究人员，我的研究方向属于化学领域，不过，很多人可能不知道，化学跟量子力学有着密切的联系。在19世纪20年代，薛定谔创立了量子力学的数学体系，其中最重要的成果就是求解了氢原子的薛定谔方程，认识了原子结构，这也成为现代化学家认识原子结构的基础。随后，化学家们就开始用量子力学处理各种原子、分子的问题，以及它们之间的反应问题，从而创立了量子化学这一学科。化学家们对于化学键以及化学反应本质的认识，就是基于量子力学得出的理论。近些年，随着量子信息科学的发展，用量子计算机来模拟化学反应成为量子计算机的一个重要应用，事实上，这也是当年费曼提出量子计算机构想的一个重要思路——用量子来模拟量子，因为化学反应在原子、分子层面都要遵循量子力学法则。费曼也曾经说过，理论化学的最终归宿是在量子力学中。所以说，量子化学其实是量子力学的一个重要分支，而我正好在讲授这门课程，因此我对于量子力学是很熟悉的。

其实早在二十年前读研究生期间，我就对量子力学非常感兴趣，这要归功于一本名为《时间之箭——揭开时间最大奥秘之科学旅程》的科普书，这本书让我体会到了量子力学的奇妙之处，直到现在，我还时常拿出来翻阅。当时，我们的“量子化学”课程由黑龙江省教学名师徐崇泉教授讲授，这门课我考了97分，得到了徐老师的赞许。后来，我到哈工大威海校区任教，讲授的第一门课就是“结构化学”，这门课是量子化学的先导课，主要就是应用量子力学原理来处理原子、分子的结构。第一次授课时，受系里邀请，徐崇泉教授专门从哈尔滨赶来对我进行指导，为我更深入地理解量子力学的内涵打下了良好的基础。此后，在一轮又一轮的授课过程中，我越来越深入地认识到了量子力学的奇妙之处，对量子力学产生了越来越浓厚的兴趣。

都说兴趣是最好的老师。任教十多年，我研读的量子物理专业书籍和科普书籍超过百本，对量子力学逐渐有了比较深刻的理解和认识，由此产生了用自己的方式把神奇的量子世界介绍给读者的想法，经过几年的打磨，我的第一本科普作品《从量子到宇宙——颠覆人类

认知的科学之旅》诞生了。感谢清华大学出版社的厚爱，感谢出版社编辑的鼓励和支持，让这本书得以出版，从此我也踏上了科普写作的道路。这本书出版以后，入选了中国图书评论学会2017年月度“中国好书”榜单，也获得了读者的广泛好评。山东省科协还专门请我去做了一次有关量子物理的讲座，为科技工作者们进行量子科普，后来我也多次为青少年做过量子科普讲座，反响都很好，这也极大地激励了我对科普创作的信心。

其实，我对上一部作品还是略感缺憾的，因为我当时专注于量子力学的理论，对于量子技术方面介绍得不够，随着近几年量子信息技术的快速发展，我很想找机会弥补这一缺憾，恰好清华大学出版社承接的出版项目中要为青少年创作基础前沿科学史丛书，其中一本就是介绍量子科学，我很高兴能承担这一重任，写一本量子力学理论与技术兼顾的科普作品。

这次创作时间紧、任务重，寒假的两个月则是我最主要的创作时间，甚至大年三十也没歇过一天，终于我自认为保质保量地完成了任务。

事实上，写科普要比讲课更困难，因为量子力学乃至量子技术的授课对象都是高年级大学生，他们已经掌握了高等数学和大学物理知识，用数学的语言跟他们对话，很多东西是比较容易讲解的，但是这次要写科普，必须假设读者是没有相关基础的，这种情况下要把量子科学和技术通俗易懂地讲解清楚，是相当困难的，这对作者的理解深度是一个极大的考验。同时，科普作品还要兼顾趣味性和可读性，既要使读者找到阅读的乐趣，也能使读者掌握基本的科学知识，还要引导读者养成科学思维的习惯，因此，优秀的科普作品其实是不多见的，我希望自己在这些方面能得到读者的认可。同时，我也希望本书能激发青少年对科学的热情，引导他们走上科学探索的道路，这也是本套丛书的创作初衷所在。最后，由于本人能力所限，疏漏和不足之处在所难免，敬请读者朋友们批评指正。

高鹏<br>2022.11

# 量子理论主要创始人获诺贝尔物理学奖一览

1918年　普朗克　能量量子化与黑体辐射的解释

1921年　爱因斯坦　光量子与光电效应的解释

1922年　玻尔　原子能量量子化与原子光谱的解释

1929年　德布罗意　实物粒子的波粒二象性

1932年　海森伯　矩阵力学理论与不确定原理

1933年　薛定谔　波动力学理论

1933年　狄拉克　狄拉克方程与量子力学理论体系

1945年　泡利　泡利不相容原理

1954年　玻恩　波函数的概率统计诠释

1965年　费曼　路径积分理论与量子电动力学

# 引言

如果你是第一次听到“量子”这个词，很有可能会以为它是某一种粒子的名字，其实不然，这是一种常见的误解。事实上，“量子”是一种物理概念，这个概念是与经典物理中的“连续”相对立的，它代表的是一种不连续的变化方式，我们称之为“量子化”。

我们所熟知的所有微观粒子，如光子、电子、质子、原子、分子等，在微观尺度里都表现出明显的量子特性，这是与我们在日常生活中的认知完全不同的特性，我们所熟悉的许多物理认知，在量子世界中都被彻底颠覆。微观粒子的运动根本不服从牛顿力学，因此，描述微观粒子运动规律的科学就被称为量子力学。

量子力学的发现过程，是一幅波澜壮阔的历史画卷，其中，既有人类智力的巅峰对决，也有超出想象的自然之谜。量子现象给人类带来的冲击和震撼，连人类最聪明的大脑都为之惊叹。量子物理对人类文明的推动作用，在过去100年已经带来了一场深刻的技术革命，且在未来的100年，还将继续带来另一场更深刻的技术革命。

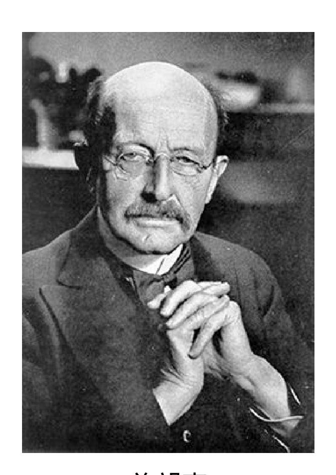
下面，就让我们跟随历史的脚步，把这幅画卷徐徐展开，跟那些伟大的物理天才们一起，去探索量子科学的奥秘吧。

# 目 录

# 第一篇　量子·起源

# 1　黑暗中的光

1900年是20世纪的第一年，从伽利略时代算起，近代物理学到这时候已经发展了近300年。300年间，物理学家们格物致理、孜孜不倦地探求自然界的奥秘，开辟出了力学、光学、热学、电磁学等多个研究领域，涌现出牛顿、法拉第、麦克斯韦、玻尔兹曼等一大批天才的物理学家。到1900年的时候，人们已经弄清楚了太阳系的运行规律，发现了元素周期表，发明出蒸汽机和发电机，甚至发明了无线电通信……人类对世界的认识和改造达到一个空前的高度，当时很多物理学家自信满满地认为，人类对自然界已经了如指掌，人类对物理学的探索也即将走到尽头，到那时候，宇宙在人类眼里将不再有秘密。



普朗克

1900年，德国物理学家马克斯·普朗克（1858—1947）刚满42岁，但他已经荣誉满身了。普朗克21岁博士毕业以后，先在自己的母校慕尼黑大学任教，后来又回到家乡的基尔大学任教。凭借自己在热力学领域的出色工作，他在1889年来到了首都柏林，出任柏林大学理论物理研究所的主任，1894年，他当选为普鲁士科学院的院士。

荣誉加身的普朗克，在世人眼里已经是一位非常成功的物理学家

了，但他自己却时常会回想起他的大学物理老师冯·约利对他说过的一番话。那时候，他一心想钻研物理，于是申请从数学系转到物理系，没想到，冯·约利居然对他说，物理学的大厦已经建成，剩下的只不过是在一些偏僻的角落里进行边边角角的修补，已经没有什么大的发展前途了。普朗克虽然没有被这些话语劝退，但是这些话却在他的心底深深地扎下了根，他也时常在疑惑，物理学难道真的快走到尽头了吗？

就在他当选院士的那一年，普朗克决定向当时物理学界的著名难题——黑体辐射发起进攻，他希望能攻克这个难题，即便是修补大厦的边边角角，他也要修补最难的那一块。

当物体被加热时，就会发光发热，例如，烧红的铁块在黑暗中会放出橙黄色的光芒（图1-1）。当时物理学家们已经知道，“光”就是电磁波，发光就是辐射电磁波，电磁波携带的能量就是测量出来的“热”。事实上，任何温度高于绝对零度（-273.15　℃）的物体都在发光发热，只不过，它们发出的“光”并非都是可见光。只有波长在400～700　nm的光才是可见光（图1-2），也就是人类肉眼能识别的电磁波，其他波段的电磁波都是不可见光，人类看不到。例如，人类虽然也在发光，发出的却是肉眼看不到的红外线。而物体只有在被加热到500℃以上时才会发出较强的可见光。



图1-1　烧红的铁块发出可见光

温度越高，辐射能力越强。热辐射看起来好像并不复杂，按道理讲，

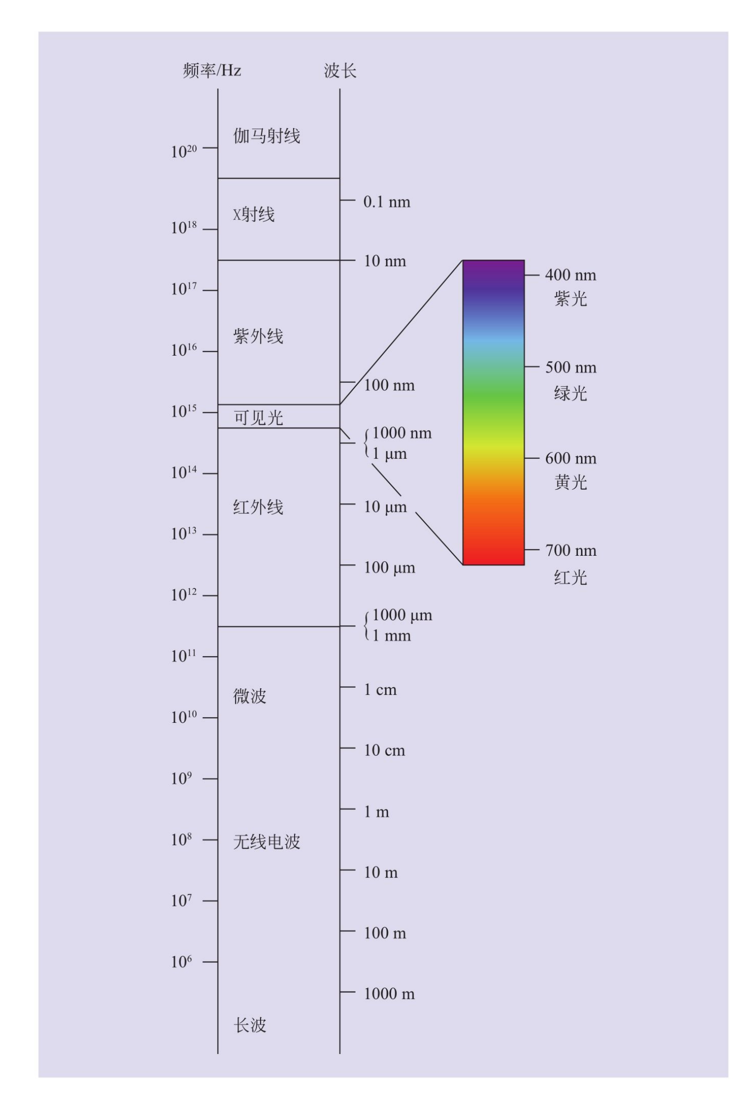



图1-2　可见光在电磁波谱中的范围

物体发光发热的现象，在物理学上有一个专有名词——热辐射。温度越高，辐射能力越强。热辐射看起来好像并不复杂，按道理讲，当时人们已经有了完善的光学、热学、统计力学、电磁学等理论，解释这个现象应该不算一个难题，但令人意外的是，这竟然是当时的一大难题。

为了研究热辐射，人们设想了一种理想情况。如果一个物体能吸收全部的外来光，那么当它被加热时就能最大限度地发光，这就是理想的热辐射，也叫黑体辐射。“黑体”的概念是普朗克的老师基尔霍夫在1862年提出来的。我们知道，一个物体之所以呈黑色，是因为它能吸光而不反光。显然，最黑的物体能把照射到它表面的所有光都吸收掉，一点儿都不反射，这就是“黑体”。

最开始人们用涂黑的铂片作为黑体来研究。后来，德国物理学家维恩想出来一个更巧妙的办法来制作黑体：找一个内壁涂黑的耐热的密闭箱子，在箱子上开一个小孔，因为射入小孔的光能被完全吸收，所以这个小孔就是一个“黑体”（图1-3）。

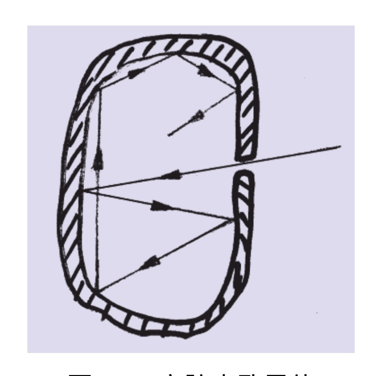


图1-3　空腔小孔黑体

当时人们通过实验已得出了黑体辐射的光波波长与辐射能量之间的关系曲线，对于一个理想的热辐射来讲，这条曲线是确定的，只随温度变化（图1-4）。但是在理论解释上，却找不到一个合适的公式来描述这条曲线。物理学家们通过经典的热力学和统计力学推导出两个公式，分别叫维恩公式和瑞利-金斯公式，但这两个公式只能分别解释曲线的一半，都无法给出全部曲线的能量密度分布。经典物理学在这个问题上，似乎无能为力。

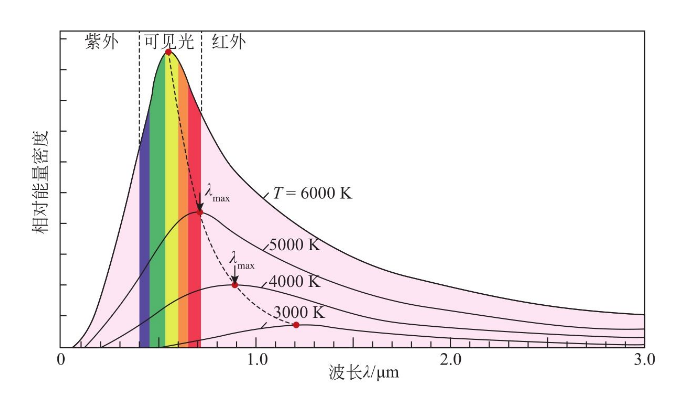



图1-4　不同温度下的黑体辐射能谱曲线

（https://www.chem17.com/tech_news/detail/2195430.html）

到1900年，普朗克研究黑体辐射问题已经6年了。身为热力学专家，顶着科学院院士的光环，奋斗6年仍然一无所获，普朗克承受的压力也是巨大的，付出和回报似乎不成比例，能否取得成果还是未知数，难道要在这个问题上耗一辈子？

耗一辈子就耗一辈子！普朗克下定决心。解决一个重大问题胜过解决10个普通问题。普朗克知道，这个问题对整个物理学至关重要。他决定，无论付出什么样的代价，都要找到黑体辐射的理论解释。

## 扩展阅读

如果我们对比一下普朗克公式和维恩公式，就会发现普朗克仅仅在维恩公式的指数项后面减了个1，这一点小小的变化，竟产生了天壤之别的结果。二者的区别如下。

维恩公式：$f(\lambda)=\dfrac{b}{\lambda^{5}e^{\frac{a}{\lambda T}}}$

普朗克公式：$f(\lambda)=\dfrac{b}{\lambda^{5}(e^{\frac{a}{\lambda T}}-1)}$

式中：$f$ (λ)是黑体辐射能量随波长λ的分布函数；$T$是温度；e是自然常数（e＝2.718…）；$a$和$b$是两个经验参数。

经过6年的研究，普朗克非常清楚，经典物理学是无法解决这个问题的。看来，必须要做出一些改变，这个改变是大是小，还不得而知，但是，必须迈出这一步。于是，普朗克决定抛弃经典物理的条条框框，先凑一个公式出来。不管公式的来由是什么，先找到一个能符合实验曲线的公式，然后再来寻找这个公式背后的物理内涵。

普朗克从维恩公式入手，结合6年来早已烂熟于心的实验曲线，经过一番推敲，最后，利用数学上的内插法，他竟然真的凑出了一个公式，这个公式可以完全解释整条黑体辐射曲线，分毫不差！这一结果让普朗克欣喜若狂，但更让他紧张焦虑，他已经看到了希望的曙光，但似乎又处在黎明前的黑暗中，他必须找到这个公式背后隐藏的物理奥秘，去迎接黎明真正地到来。

接下来的几个星期，是普朗克一生中最忙碌最紧张的几个星期，他的全部心思都花在了这个公式上面，他不满足于仅仅出于凑巧找到这个公式，他的目标是把这个公式推导出来。他的大脑不停地高速运转，日夜推算这个公式背后的秘密，渐渐地，一幅完全意想不到的图景在他的脑海中清晰起来——能量可以是不连续的吗？他不断地问自己。

在经典物理学中从来没有人问过这个问题，或者说从来没有人意识到这是一个问题。所有人都下意识地认为能量一定是连续的，就像我们在数学中处理一条光滑的曲线一样，可以取到曲线上任意一点的值。但是，普朗克脑海中的图景却不断地告诉他，要想把这个公式推导出来，能量就必须不连续！最终，普朗克痛苦地做出决断，接受能量的不连续性，不管这和经典物理是多么格格不入。

1900年12月14日，在柏林科学院的会议上，普朗克宣读了题为《黑体光谱中的能量分布》的论文，在这篇论文中，他提出了石破天惊的能量量子化假设：电磁辐射的能量不是连续的，而是一份一份的。他将这一份一份的能量单元称为“能量量子”。从此，量子理论正式诞生了。

在普朗克的假设里，就像物质是由一个个原子组成的一样，电磁波的能量其实也是由一份份能量量子组成，每个能量量子携带的能量

可以用一个简单的公式表示：

$$E = h\upsilon$$

其中：υ是电磁波频率[(1)]；$h$是普朗克提出的一个新的物理学常数，叫做普朗克常数（$h \approx 6.262 \times 10^{-34}$ J·s）。

能量量子化的概念，是一个全新的、从未有人想到过的概念，经典物理学的大厦里，根本没有这个概念的容身之处。普朗克的老师认为物理学的大厦即将完成，但是，也许普朗克自己都没有意识到，他已经为一座新的大厦的奠基铲起了第一锹土，造出了第一块砖，这座新的物理学大厦就叫量子力学。

量子力学这个名词是和经典力学相对应的，经典力学就是牛顿力学，它研究的是宏观世界里物体的运动规律，而量子力学研究的则是微观世界里粒子的运动规律。宏观和微观的分界线，就取决于普朗克常数。

普朗克常数是量子力学的标志性常数，可以反映微观系统的空间尺度、能量量子化特征等，因此它也成为界定经典物理与量子力学适用范围的重要参数。当普朗克常数的影响趋于零时，量子力学问题将会退化成经典物理问题。由于普朗克常数非常非常小（图1-5），因此，它对宏观物体和宏观运动的影响基本上等于零，这也是我们在日常生活中看不到量子效应的原因，所以人们才一直误以为能量是连续的。也幸亏普朗克常数如此之小，才让我们的日常世界井然有序、有章可循，如果你进入量子世界，那里变幻莫测的混乱景象可能会使你彻底晕头转向、再无章法可依。当然，这一点，当时的物理学家们还都不知道，普朗克只是造出了第一块砖，量子力学的大厦，还需要更多的天才物理学家们一点一点地构筑。

## 扩展阅读

在物理学的发展过程中，每当一个重大理论被提出的时候，总是有一个相应的标志性的普适常数出现。例如，牛顿力学中的引力常量、热力学与统计物理中的玻尔兹曼常量、相对论中的真空光

速，乃至量子力学中的普朗克常数。

这些常数不仅是相应理论的标志，而且也能反映出各理论之间的关系。例如，物体的运动速率与光速的大小关系成为判断牛顿力学适用范围的一个重要参照，只有当物体速度远远小于光速的时候，牛顿力学才是适用的；或者说，只有当物体速度接近于光速的时候，相对论效应才变得明显。

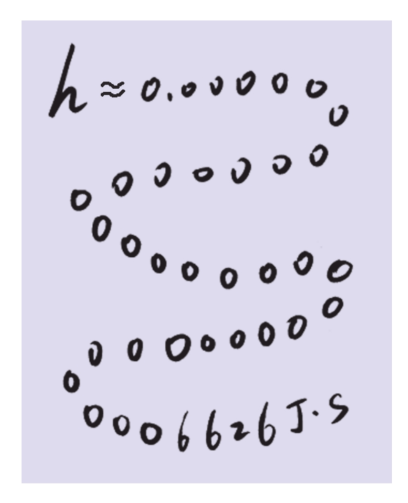


图1-5　普朗克常数



什么是量子？为什么说量子化才是世界的本质？

---

(1)υ为希腊字母，读音为/nju:/，相对应的另一个表示波长的字母λ读音为/'læmdə/。

# 2 光雨

1905年是物理学史上非常重要的一年，这一年诞生的理论，奠定了整个20世纪物理学的基础，而这些所有的理论竟然都是由同一个人提出来的，他就是阿尔伯特·爱因斯坦（1879—1955）。

1905年，爱因斯坦连续发表了4篇论文。这4篇论文每一篇都具有划时代的意义——第一篇解释了光电效应，提出光子的概念，是量子理论的重大发展；第二篇解释了布朗运动，提供了原子存在的重要证明；第三篇提出了狭义相对论，相对论正式诞生；第四篇揭示了质能关系的深层本质，质能方程$E=mc^2$以其简洁优美的形式风靡全世界，成为相对论的代名词。后来，1905年被称为“爱因斯坦奇迹年”。



爱因斯坦

这一年，爱因斯坦刚刚26岁。这一年，距离普朗克提出能量量子化的观点已经过去了5年，但是在这5年中，量子理论没有任何发展，欧洲各所大学的知名教授们，都还在忙忙碌碌地修补着经典物理学的大厦，没有人能意识到能量量子化到底意味着什么，普朗克的工作几乎无人问津。连普朗克自己都陷入了深深的自我怀疑当中，他对黑体辐射公式的推导存在严重的内在矛盾，这让他觉得能量量子化也许只

是权宜之计，难登大雅之堂，所以他一直在尝试如何才能重新回到经典物理学的框架中去推导黑体辐射公式。

在这一年之前，谁也不会想到，全欧洲最有才华的物理天才竟然是瑞士专利局的一个小职员。此时的爱因斯坦，没有加入任何学术组织，只与几位热爱科学与哲学的好友组织了一个叫做“奥林匹亚科学院”的读书俱乐部。几个年轻人都不是学术圈的人，他们有日常养家糊口的工作要做，但从“奥林匹亚科学院”这个颇有气魄的名字就能看出，这是一群志向远大的年轻人。他们挤出周末或者下班时间聚在一起，就他们感兴趣的话题——哲学、物理、数学和文学——一边读书一边讨论。

爱因斯坦的学术之路之所以从专利局起家，并不是他不愿意步入学术殿堂，而是没有一所大学能接纳他。爱因斯坦上大学的时候经常逃课，给老师留下了很差的印象。他不是不爱学习，而是认为老师讲的东西都过时了，无法满足自己的需求，于是就逃课躲到外面去自学。他通读了基尔霍夫、赫兹、玻尔兹曼、洛伦兹、麦克斯韦等物理大师的著作，了解了物理学最前沿的内容，但是这对他的毕业考试并没有太大的帮助，毕竟老师考的重点不在他的阅读范围之内，这导致他的毕业成绩不佳。1900年，也就是普朗克提出能量量子化的那一年，爱因斯坦大学毕业，当时他一心想留校做助教，但是他的老师理所当然地拒绝了一个总是逃课的学生。然后他又给欧洲各所大学乃至中学发出了求职信，但都没有回应。蹉跎两年之后，他才在大学好友的帮助下找到了专利局技术员这样一份工作，总算没有沦落为一个无业青年。

工作和生活稳定下来以后，爱因斯坦终于不用再为养家糊口发愁了，他可以静下心来，研究他心爱的物理学了，纵使只能在业余时间做研究，但对他来说也已经是很难得了。他始终保持着敏锐的目光，追踪着物理学的前沿进展，对物理学的各个方向都有所研究。

“光”是爱因斯坦始终关注的一个焦点，无论是光的速度还是光的本性，都是他思考的问题。这期间，他既了解到普朗克对于黑体辐射问题的解决，也在思考着另一个奇怪的与光有关的难题——光电效应。

1887年，德国物理学家赫兹通过实验首次证实了电磁波的存在，随后，他又证明了光波就是电磁波，全面验证了麦克斯韦的电磁理论。但是，赫兹在验证经典电磁理论的同时，还发现了一个异常的实验现象——光电效应。

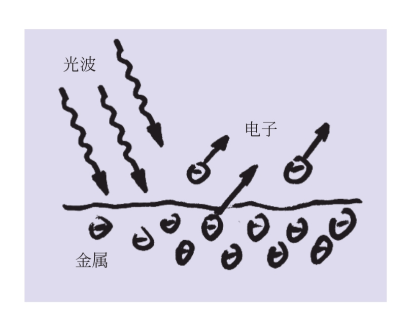



图2-1　光电效应示意图

光电效应，顾名思义，就是由光产生电的效应（图2-1）。金属是由原子构成的，原子又是由原子核和电子组成的。赫兹发现，用紫外线照射某些金属板，可以将金属中的电子打出来，在两个相对的金属板上施加电压，被打出来的电子就会形成电流。这一现象引起众多研究者的兴趣，很快就得到了大量的实验结果，可是电磁波理论在解释这些实验结果时却遇到了严重的困难。

人们发现，决定能否打出电子的关键，不在于光的强度，而在于光的频率。紫外线可以轻易从金属中打出电子，而可见光却不行。当时人们对此百思不得其解，因为按照经典的波动理论，波的强度便代表了它的能量，只要光强足够，就能使电子获得足够的能量脱离金属表面的束缚，所以应该任何频率的光都能打出电子，可实验结果却是再强的可见光也打不出电子，与理论预测完全相反。

自光电效应被发现以来，已经过去了将近20年，但是这一难题仍然无人能解。正所谓初生牛犊不怕虎，面对这样公认的科学难题，年轻的爱因斯坦并没有畏缩，他敏锐的直觉告诉他，经典的电磁理论主要描述宏观上光的整体性质，而黑体辐射和光电效应本质上都涉及微观上光的产生过程，既然普朗克通过能量量子化解决了黑体辐射问题，那么光电效应问题应该也可以从中获得启发。

为什么光电效应中光的频率这么重要呢？爱因斯坦紧紧盯着普朗克的能量量子公式：

$$E=h\upsilon$$

从这个公式来看，能量量子携带的能量只与光的频率$\upsilon$有关，当光照射到金属表面时，其实就是能量量子在不断地冲击金属表面，那么能量量子到底表示什么呢？是一小段波？是最微小的振动？还是别的什么？从普朗克的论文来看，普朗克并没有给出能量量子的明确图像，而这幅图像，应该是至关重要的。

陷入沉思的爱因斯坦，仿佛老僧入定了一般，一动不动，没有人能看得出来，他那天才的大脑正在高速运转。渐渐地，他的眼前仿佛出现了一幅画面：光的能量量子就像一颗颗子弹射向金属内部，被子弹击中的电子获得了子弹的能量，便从金属内部的束缚中挣脱出来。

有了！爱因斯坦一拍桌子，猛地站起身来，电子能不能被打出来，就完全取决于子弹到底能给电子提供多少能量！他激动地在屋子里转了几圈，然后坐在桌子前拿起草稿纸，赶紧推演起来，很快，一个公式就跃然纸上：

$$\text{电子动能}= h\upsilon\text{-电子逸出功}$$

这个公式的意思是：能量量子给电子提供了大小为$h\upsilon$的能量，这些能量除了要帮助电子挣脱金属表面的束缚外（电子逸出功），剩下的就变成了电子的动能。

至此，能量量子的图像在爱因斯坦的头脑中已经完全明确了，这一小份一小份的电磁辐射能量并不是一小段一小段的波，而是一个个粒子，这些粒子是不可分割的，只能被整个的吸收或者发射。爱因斯坦给这些能量点粒子起名为光量子，后来人们改称为光子。

爱因斯坦的光子理论很好地解释了光电效应。因为每一个光子的能量都是固定的$h\upsilon$，那么光照射到金属表面，电子所吸收的能量主要取决于单个光子的能量而不是光的强度，光的强度只是光子流的密度而已。因为可见光频率低，其光子的能量不够大，不足以克服电子逸出功，所以没法打出电子。而紫外线频率高，光子能量大，所以很容

易打出电子。

爱因斯坦提出光子假设是很大胆的，因为当时还没有足够的实验事实来支持他的理论。直到1916年，才有美国物理学家密立根对他的理论作出了全面的验证。有趣的是，密里根在做光电效应实验时，本来是想推翻爱因斯坦的光子理论，所以他一直做了10年的实验，10年间，他不断地提高实验的精度，结果却发现实验精度越高，越能证明爱因斯坦的正确性，最后没办法，他只好承认了爱因斯坦的理论，而且还顺便比较精确地测定出了普朗克常数的值。

爱因斯坦在明确了光具有粒子性以后，随后又进一步根据相对论提出了光子的动量公式：

$$p=h/\lambda$$

式中，$p$为光子的动量；$\lambda$为光的波长；$h$为普朗克常数。

1923年，美国物理学家康普顿和他的学生吴有训通过康普顿效应的验证实验，证实了光子的确具有动量，为光具有粒子性提供了无可辩驳的证据。

关于光的本性，在历史上曾经有过长期的争论。在17世纪末，以牛顿为代表的粒子派和以惠更斯为代表的波动派进行过长期论战。18世纪，人们发现了光的干涉、衍射等现象，波动说全面占了上风。19世纪后期，随着电磁理论的问世，人们明确了光就是电磁波，粒子论被彻底抛弃。结果没过多少年，爱因斯坦又重新提出光子学说，明确了光的粒子性，那么这一次，光的波动性又该如何看待？

事实上，爱因斯坦在光子理论的两个公式中已经给出了答案。我们再来看一下这两个公式：

$$E=h\upsilon\text{（光子能量＝普朗克常数×光的频率）}$$

$$p=h/\lambda\text{（光子动量＝普朗克常数/光的波长）}$$

这两个公式看起来简单，实际很不简单。因为爱因斯坦通过这两个公式把粒子和波联系起来了：粒子的能量和动量是通过波的频率和波长来计算的。也就是说，爱因斯坦把光同时赋予了粒子和波的属性，光具有波粒二象性！

波粒二象性的发现，是人类对光的本质的认识的重大突破，由此带来的“蝴蝶效应”，将使人类对物质世界的认识发生重大飞跃，这一飞跃，将在18年后由法国科学家德布罗意做出。而此时，德布罗意还只是一个13岁的小男孩，他正沉浸在历史和文学的海洋中，立志将来要做一名历史学家。

# 3 文科生的逆袭

1911年对于国际物理学界来说是一个重要的年份，因为这一年是历史上最负盛名的物理学术盛会——索尔维会议首次召开会议的年份。参加这次会议的物理学家，大多数都是出现在当今物理教科书里的人物，其阵容之豪华，堪比武侠小说里的华山论剑（图3-1）。

这次会议是在比利时首都布鲁塞尔举办的，由比利时的化学家兼实业家索尔维赞助。这次大会的主题是“辐射与量子”，由德高望重的荷兰物理学家洛伦兹主持，专门讨论刚刚登台的量子论。显然，物理学家们已经意识到了量子论对于经典物理学的冲击，他们必须做一次深入的交流与讨论，以把握未来物理学的走向。但是，大多数科学家显然还没做好准备迎接新时代的到来，第一个做报告的是洛伦兹，他用德语、法语和英语三种语言轮流讲演，讲得极为精彩，但是，他演讲的题目却是“用经典的方法讨论辐射问题”。

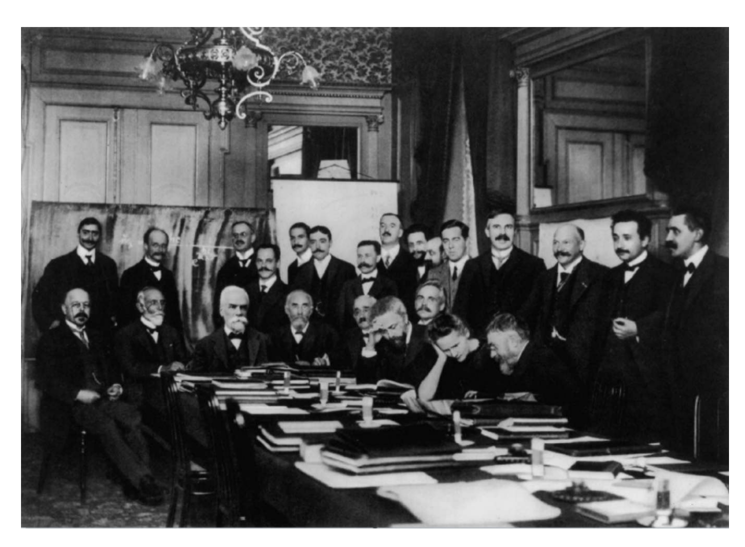


图3-1　第一次索尔维会议合影

坐者（从左至右)：（1）沃尔特·能斯特；（2）马塞尔·布里渊；（3）欧内斯特·索尔维；（4）亨德里克·洛伦兹；（5）埃米尔·沃伯格；（6）让·佩兰；（7）威廉·维恩；（8）玛丽·居里；（9）亨利·庞加莱。站者（从左至右）：（1）罗伯特·古德施密特；

（2）马克斯·普朗克；（3）海因里希·鲁本斯；（4）阿诺·索末菲；（5）弗雷德里克·林德曼；（6）莫里斯·德布罗意；（7）马丁·努森；（8）弗里德里希·哈泽内尔；（9）豪斯特莱；（10）爱德华·赫尔岑；（11）詹姆斯·金斯；（12）欧内斯特·卢瑟福；（13）海克·卡末林·昂内斯；（14）阿尔伯特·爱因斯坦；（15）保罗·朗之万。

普朗克和爱因斯坦都参加了这次会议，这也是两位巨星的首次会面。此时的爱因斯坦已经是布拉格大学的理论物理教授了。这次会议上，爱因斯坦终于说服了普朗克接受他的光量子理论。要知道，在这之前，普朗克对光量子是持反对态度的。事实上，参加这次会议的大多数科学家都是反对光量子理论的，他们大都希望维持经典物理学的体系，唯独居里夫人是个例外，她坚定地支持爱因斯坦。虽然这是她第一次见到爱因斯坦，但她独具慧眼，对爱因斯坦那透彻的分析能力极为欣赏。

对爱因斯坦来说，这次会议乏善可陈，除了和普朗克、居里夫人、朗之万等几位科学家结下了深厚的友谊之外，面对经典物理学的顽强反抗，他也无可奈何。事后，他对这次会议的总结是：“啥也没讨论出来。”

尽管当时“啥也没讨论出来”，但是，这次会议后来却取得了一项“重大成果”——吸引了一位学历史的青年学生改行攻读物理学位，这个青年人就是路易·德布罗意（1892—1987）。

第一次索尔维会议的秘书是法国物理学家莫里斯·德布罗意，他是研究X射线的专家，也是路易·德布罗意的哥哥。莫里斯回家以后，把这次会议的见闻以及这些著名人物的辩论兴致勃勃地给自己的弟弟讲述了一番，还把会议资料拿给弟弟看。

路易·德布罗意本来是学历史的，但是他哥哥在家里建了一座实验室，耳濡目染之下，他对物理也有所了解。这一次，他哥哥讲述的会议见闻让他对这些物理大师向往不已，会议资料里爱因斯坦和普朗克有关量子化概念的文章也让他产生了极大的兴趣，于是，这位历史专业的学生决定放弃历史，转攻物理。

两年后，德布罗意拿到了理学学士学位，这时候，第一次世界大战爆发，德布罗意被征召入伍，在巴黎的埃菲尔铁塔军用无线电报站

服役了6年。1918年年底，“一战”结束，德布罗意随后退役。1919年，他回到巴黎大学跟随朗之万攻读物理学博士学位。



德布罗意

博士生的研究工作需要靠自己独立完成，导师只是提供一些参考意见，于是，德布罗意决定研究自己最感兴趣的量子理论。

从1911年到1919年，短短8年间，随着实验证据的不断出现，物理学界对爱因斯坦的光量子理论已经从普遍反对变成了普遍接受。爱因斯坦提出的光量子理论把原来不相干的波和粒子糅合在了一起，体现出波粒二象性的特点。但是，因为光是一个很特殊的东西，光子的静止质量为零，光速又是所有速度的极限，所以大家也能接受光具有波粒二象性这样的特殊性质，并没有考虑这个性质是否具有普遍性。

德布罗意一直在认真思考光的波粒二象性，他隐约觉得这个现象并不简单，背后或许隐藏着一些更深层次的奥秘，那会是什么呢？他日夜苦思冥想。

时间一晃到了1923年。有一天，一丝亮光突然出现在他的脑海中，于是德布罗意灵光乍现、顿悟天机：既然一度被视为波的光具有粒子性，那么反过来，一直被认为是粒子的物质粒子会不会也具有波动性呢？

正所谓厚积薄发，灵感一旦到来，他的思路豁然开朗。德布罗意立刻意识到，波粒二象性应该具有普遍性，爱因斯坦1905年的发现应当得到推广，运用到所有的物质粒子，特别是电子上。博士论文的课题，有了！

当然，德布罗意的观点并不是泛泛的哲学观点，他在博士论文里面展开了大量的定量讨论。经过近一年的努力，德布罗意在1924年完成了他的博士论文——《量子理论研究》。在论文中，德布罗意把爱因斯坦的公式原封不动地搬运过来，指出实物粒子在运动时，伴随着波长为λ的波，粒子的能量和动量与波的频率和波长有以下关系：

粒子能量$E=h\upsilon$

粒子动量$p=h/\lambda$

后来，人们把这种波叫德布罗意波，也叫物质波。

德布罗意是采用类比的方法提出他的假设的，当时并没有任何直接的实验证据，所以，当他参加博士论文答辩的时候，在场的专家问他："如何用实验来证实你的理论呢？"

对于这个问题，德布罗意早就准备好了，他知道，答辩时一定会有人提出这个问题，所以他早就想好了回答："如果让电子通过晶体，它应该会产生一个可观测的衍射现象，这样就能证明它的波动性！"

但是，德布罗意自信满满的回答，并没有完全打动评委。他的导师朗之万和在场的4位评委都对这个大胆的假设充满疑虑，如果授予他博士学位，万一贻笑大方，导师也跟着丢人；可是如果不授予他学位，万一他的想法是正确的，岂不是误人终身？

面对这样艰难的抉择，朗之万想到了爱因斯坦，他们在第一次索尔维会议上结识以后就成了好朋友，经常通信联系。于是，朗之万决定暂缓公布结果，把论文寄给爱因斯坦，听听爱因斯坦的评价再做决定。

很快，朗之万就收到了爱因斯坦的回信。爱因斯坦不愧是爱因斯坦，他具有非凡的科学洞察力，他在回信中对论文给予了极高的评

价，并写道：“德布罗意揭开了物理学厚重大幕的一角。”这下子，德布罗意的博士学位终于稳了。

## 扩展阅读

波的一个特性是遇到障碍物（如狭缝、小孔等）后会绕过其传播，这就是衍射。只有在障碍物的大小与光的波长接近时，才能观察到衍射现象。

如果在一块平板上制作一系列极窄的平行狭缝，就构成了一个光栅，可以用来观察光的衍射现象。

按照德布罗意波的公式计算，实物粒子的波长是非常小的。例如，电子在1000 V的加速电压下，波长仅为39 pm，波长的数量级和X射线相近，所以用普通光栅很难检验其波动性。不过晶体倒是一种天然的光栅，因为晶体中原子有序排列可以形成晶面，同一方向晶面平行等距排列，类似于一系列平行狭缝，且“狭缝”间距与电子波长相近，因此可以用来检验电子的波动性。

爱因斯坦认识到，德布罗意的发现具有重大意义，应该尽快将其成果向学术界推荐。几个星期后，爱因斯坦就在自己撰写的一篇论文中专门介绍了德布罗意的工作。他写道：“一个物质粒子可以怎样用一个波场相对应，德布罗意先生已在一篇很值得注意的论文中指出了。”

爱因斯坦的推荐立刻引起了科学界的重视，实验物理学家们开始寻找物质波。1927年，英国科学家G.P.汤姆孙让电子穿过金箔，果然得到了电子的衍射图像（图3-2），而且波长与计算结果一致，证实了德布罗意波的存在。需要注意的是，一个电子在屏幕上只能打出一个亮点，电子衍射图像是由一个个电子的落点重叠起来而显现出来的，这是德布罗意波与经典波的一个重要区别，其物理内涵我们将在后面介绍。

实物粒子波粒二象性的发现，是量子力学史上一个重要的里程碑

式的事件，量子力学的大厦，终于打好了地基，快要拔地而起了。

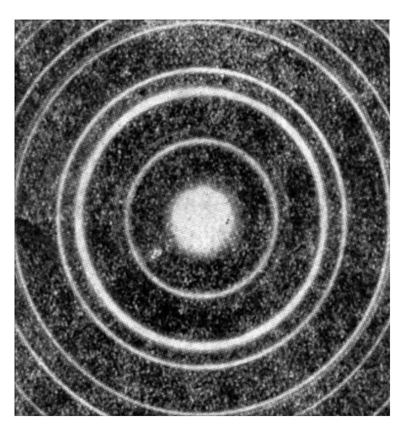


图3-2　电子衍射图像（样品为金箔）

# 第二篇　量子·创立

# 4　指纹玄机

著名的量子物理大师费曼编写的物理教科书《费曼物理学讲义》是风靡全世界的物理学经典教材，这部教材的开篇第一章就介绍了原子的运动。关于原子的重要性，费曼在书中写道："假如有一天由于某种大灾难，人类所有的科学知识都丢失了，只能有一句话传给下一代，那么怎样才能用最少的词汇来传达最多的信息呢？我相信这句话是原子的假设：所有的物体都是由原子构成的——这些原子是一些小小的粒子，它们一直不停地运动着，当彼此略微离开时相互吸引，当彼此过于挤紧时又互相排斥。"

原子是如此重要，所以对于科学家们来讲，搞清楚原子的结构是了解物质世界的基础。现在，我们在量子力学的帮助下，已经对原子的结构有了比较清楚的认识，但是，在1925年之前，物理学家们还处在迷茫之中，那时候，原子的结构还是一个深奥的科学难题。

我们周围的物质都是由原子构成的，原子又是由带正电荷的原子核和带负电荷的电子构成的（图4-1）。原子的半径只有0.1　nm大小，一滴水里就包含了大约10万亿亿个原子。打个比方来说，如果把一个网球里的原子放大到网球那么大，那么这个网球就会变得像地球一样大！原子这么小，是很难被人看到的，所以在历史上关于原子是否存在曾经有过激烈的争论。好在，现在科学家们借助电子显微镜已经直接观察到了原子，这已经是确定无疑的事实。

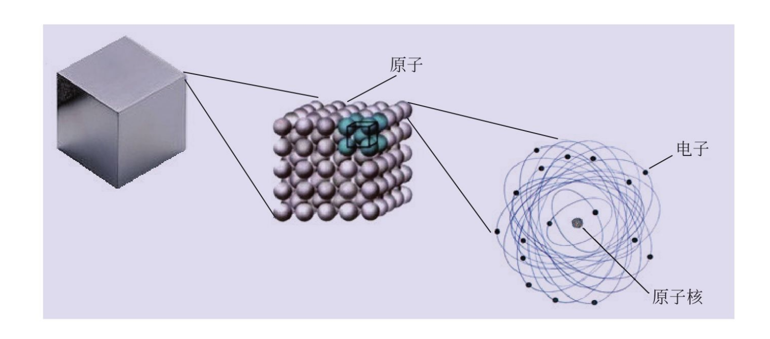


图4-1　物质是由原子构成的

原子核在原子的中心，它占据了整个原子质量的99.99%以上，而原子核的体积却非常非常小。即使把原子放大到一个足球场那么大，原子核也只有绿豆那么小！电子在原子核周围运动，电子更是小得几乎没有体积。也就是说，原子内部大部分地方都是空的。

由于原子实在是太小了，所以电子在核外到底如何运动就只能靠猜。当然，猜也不是乱猜，构建一个原子模型以后，必须能解释已有的实验现象，这样才说明我们的猜测是有道理的。所以，一个合理的原子模型必须要能解释一个很早就被发现的实验现象——原子光谱。

大家都知道牛顿用三棱镜分光的实验，太阳光可以被分解为赤橙黄绿青蓝紫这样连续的光谱。但是在19世纪中期，人们发现并非所有的光谱都是连续的，原子的光谱就不连续。人们发现，将物质气化转变成气态原子，这时候再经过分光仪分光后，得到的光谱是一条条特定波长的分离的谱线，而且每一种元素的光谱都不一样（图4-2）。

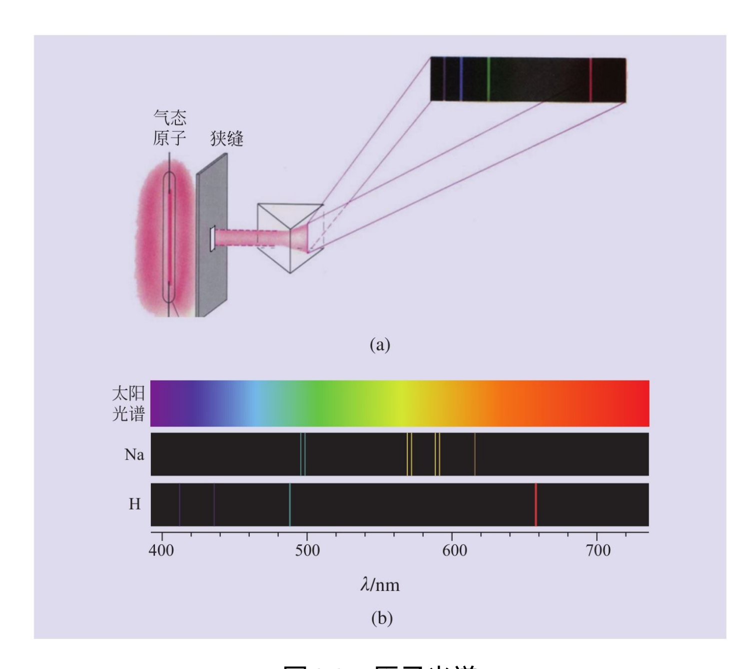



图4-2　原子光谱

（a）原子发射光谱的测试原理；（b）太阳的连续光谱和原子的线状光谱

原子光谱就像人的指纹一样，可以用来鉴定元素。在那个年代，通过原子光谱来确认新元素的发现是常用的手段，例如，居里夫人发现镭元素就是通过光谱鉴定出来的。但是，令人尴尬的是，虽然这一技术早已广泛应用，其背后的原理却还没有搞清楚，人们弄不明白为什么原子光谱是特定的谱线而不是连续的光谱。

## 扩展阅读

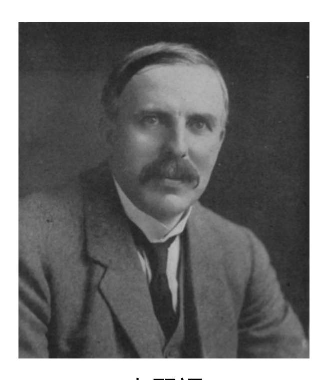
巴耳末找到的公式如下：

$$\frac{1}{\lambda}=\left(\frac{1}{2^{2}}-\frac{1}{n^{2}}\right)\times \text{常数}\ (n=3, 4, 5, 6)$$

其中，λ是当时发现的氢的前4根谱线的波长。




经典物理学在这个问题上是无能为力的，虽然瑞士的一位中学数学教师巴耳末在1885年找到了氢原子谱线的一些规律，但是他所找到的规律完全是依靠数学直觉，就像我们平时在一堆杂乱的数字中寻找规律一样，其中没有任何物理依据，没有人知道为什么会有这样的规律，也没法从理论上给予解释。



卢瑟福

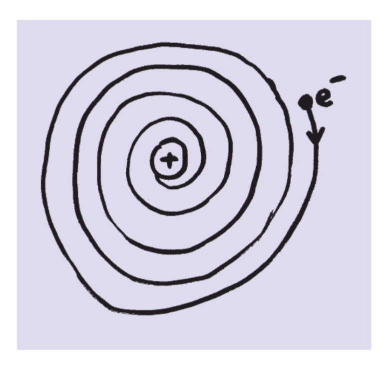
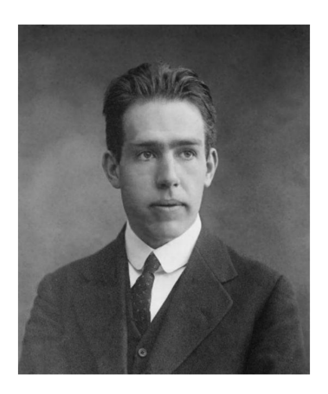



图4-3　卢瑟福原子模型面临的塌陷问题

1911年，英国物理学家欧内斯特·卢瑟福（1871—1937）发现了原子核，并提出了原子的太阳系模型。他把原子类比为一个微型的太阳系，电子被带正电的原子核吸引，围绕原子核进行轨道运动，就像行星围绕太阳运行一样。这个模型看起来很美好，宇宙中的极小（原子）和极大（星系）有着相似的运行规律，显示出自然界的和谐。但是，理想很美好，现实很残酷，这个模型存在巨大的困难，按经典电磁理论，电子在绕核运动的途中会释放能量，轨道也会逐渐变小，最后掉到原子核里，原子转瞬之间就会毁于一旦（图4-3）。但事实上这一切都没有发生，物质世界运行得井井有条，这只能说明这个模型存在着巨大的缺陷。

1912年4月，27岁的丹麦物理学家尼尔斯·玻尔（1885—1962）来到卢瑟福的实验室访学。玻尔在一年前刚刚博士毕业，并到英国剑桥大学去访学，但是在剑桥大学他没有找到合适的研究方向，恰好卢瑟福去剑桥大学做讲座，讲了他新提出的原子结构太阳系模型，玻尔立刻被这个迷人的模型吸引住了，于是追随到卢瑟福门下来求学。



玻尔

虽然在卢瑟福门下访学只有4个月，但这4个月让玻尔详细地了解了卢瑟福模型的结构以及其中的疑难。回到丹麦后，他继续潜心研究，希望破解原子结构的奥秘。这时候，玻尔已经有了用量子理论来解释原子结构的想法，但是还没有一个清晰的思路。有一次，他和同事闲聊的时候，同事建议他把原子模型和氢原子光谱联系起来考虑，并让他关注一下巴耳末公式。正所谓一语点醒梦中人，当玻尔一看到巴耳末公式，他一下子就把原子谱线和能量量子化对应起来了，一幅物理图景在脑海中悄然浮现，一切都再清楚不过了。

1913年，玻尔终于成功了，他引入能量量子化和光量子的观点，指出原子轨道的能量是量子化的，电子只能存在于能级不同的分立的轨道上。这样，电子的能量变化只能从一个能级突变到另一个能级，这个变化过程是不连续的，是突跃式的，没有中间的过渡状态，所以叫做跃迁。电子在不同轨道能级之间跃迁的时候，能量变化是固定的，而且能量是以光子形式辐射或吸收的，光子的能量为

$$\Delta E = h\upsilon$$

式中，$\Delta E$是两个跃迁轨道的能量之差，也就是光子的能量；$\upsilon$为光子的频率。由于不同轨道的能级差是固定的，于是就只能发出特定波长的光子，形成分离的谱线，如图4-4所示。

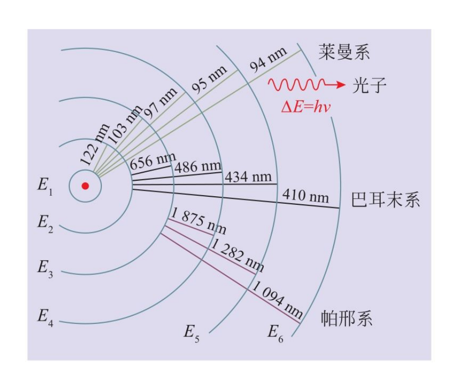



图4-4　玻尔原子模型解释氢原子光谱示意图，电子在各个轨道能级间跃迁会吸收或放出不同波长的光子，从而形成原子光谱

玻尔利用量子理论成功地解释了氢原子光谱，揭示了原子的结构，从而一跃成为量子领军人物。但是，他的模型也有明显的缺点。例如，原子轨道的能量量子化只是作为人为规定放在那儿，显得太过生硬，另外，它也只能解释氢原子的光谱，对其他原子的光谱则会出现很大的偏差。所以，原子的奥秘还没有被真正的揭示，还需要等待量子理论的进一步发展。

这一晃就是十几年，时间很快来到了1925年。

在爱因斯坦的大力推动下，德布罗意关于物质粒子波粒二象性的工作引起了物理学界的普遍关注。1925年年底，苏黎世大学物理系主任德拜听说了这一消息，他知道本系教授埃尔温·薛定谔（1887—1961）正在做量子统计方面的研究，熟悉量子领域，就请他为大家做一次报告，将德布罗意的物质波理论介绍给全系教师。



德拜没有找错人，薛定谔当时已经了解了德布罗意的工作。那段时间，薛定谔为了研究玻色-爱因斯坦统计（参见第8章），曾经多次与爱因斯坦通信进行讨论，并且从爱因斯坦的论文中了解到了德布罗意波。他在1925年11月3日写给爱因斯坦的信中说：“几天前我怀着最大的兴趣阅读了德布罗意富有独创性的论文，并最终掌握了它。我是从您那关于简并气体的第二篇论文的第8节中第一次了解到它的。”

薛定谔是一个严谨而认真的人，为了这次报告会，他重新研读了德布罗意的论文，弄清了每一个细节。果然，到了汇报那天，他做了一个清晰而漂亮的报告，自己颇为满意。但是，德拜听完之后，却不屑地点评道：“讨论波动而没有一个波动方程，太幼稚了。”

言者无心，听者有意，一句话点醒了薛定谔。薛定谔意识到，这的确是德布罗意学说的不足之处，但这同时正是自己建功立业的好机会，他马上投入到了波动方程的寻找中。几个星期后，他就成功了，这个方程也自然被命名为“薛定谔方程”。薛定谔首先为物质波定义了一个波函数，然后通过薛定谔方程描述波函数随时间的演化过程，由此可以获知量子体系的状态变化。他很快就发表了几篇论文，一举成为量子力学的奠基人。薛定谔的理论以薛定谔方程为核心，用波函数描述物质波，所以被人们称为波动力学。

事实上，薛定谔方程并不是从理论上推导出来的，而是作为假设提出来的。凭借深厚的数学和物理功底，薛定谔从经典力学和几何光学的对比入手，分析物质波的波动方程应该具有的特点，从而提出了薛定谔方程。那么，人们凭什么相信它呢？关键就在于，薛定谔在论文中建立了氢原子的薛定谔方程并求解，求解结果与原子光谱实验测定值吻合得非常好，而且也为玻尔模型中生硬的能量量子化假设找到了理论依据——量子化的得出是由薛定谔方程“自然地”求解得到的，而不像玻尔那样是人为“强加”给粒子的，这样对能量量子化的解释就更为合理和顺畅。

## 扩展阅读

求解氢原子的薛定谔方程，可得到电子的能量为

$$E_n=-\frac{1}{n^2}\times 13.6\ \text{eV}\ (n=1, 2, 3,\cdots)$$

式中，$n$是在求解过程中自然引入的参数，只能取正整数，称为量子数。eV叫做“电子伏特”，是一种很小的能量单位，表示1个电子通过1伏电压加速后所获得的能量。

因为将电子离核无穷远时的势能定为0，所以电子能量都是负值。可以看出，由于$n$只能取正整数，所以电子的能量只能取-13.6 eV、-3.4 eV、-1.51 eV等这样离散的数值，而不可能是别的数值，这就说明它的能量是量子化的。

$n=\infty$…………$E=0$

$n=4$————$E=$-0.85 eV

$n=3$————$E=$-1.51 eV

$n=2$————$E=$-3.4 eV

$n=1$————$E=$-13.6 eV

在玻尔的原子模型中，电子像“行星绕日”一样在环形轨道上运行，这是一种假想，并没有科学依据；而在薛定谔的模型中，通过求解薛定谔方程得到的波函数来描述电子的运动状态，更为科学。虽然二者是明显不同的，但是为了方便，人们仍然沿用了当初“轨道”的叫法，把电子的波函数称为“原子轨道”。那么，波函数是如何描述电子的运动状态的呢？这一点，需要等到波函数的物理意义被真正揭示以后才能水落石出。



量子力学是如何诞生的?

# 5　原来是骰子

科学的发展从来都不是一帆风顺的，薛定谔虽然找到了物质波的波动方程，并用波函数来描述物质波，获得了巨大的成功，但是关于波函数的物理意义，却在物理学界引起了激烈的争论。



图5-1　水面波

波是人们很早就注意到的一种现象，将石子投入水中，水面会上下起伏，发生振动，振动由近及远向四周水面扩散，就形成了水面波（图5-1）。敲钟时，撞击引起周围空气的振动，此振动在空气中不断传播，就形成声波。于是，人们就把以一定速度传播的振动叫做波。经典的波动是机械波，它需要传播介质，可以扩散和消失，会在空间中弥散开来。

我们觉得机械波很好理解，是因为它是一种真实的波动，物理学家们用波函数来描述机械波中介质质点振动时的位移变化规律。

为了描述物质波，薛定谔也要为其找一个波函数，于是他提出一条基本假设：一个粒子的运动状态可以用一个坐标波函数$\psi(x, y, z, t)$来描述[1]。

之所以称作坐标波函数，是因为波函数的数值是随着坐标变化的，不同坐标点的数值是不一样的。显然，在某一时刻$t$下，在空间中

每一点（$x$，$y$，$z$）上波函数都有一个数值，也就是说，波函数表示的是粒子在空间中的一种存在状态，所以现在人们更愿意称之为态函数。但薛定谔坚持认为，他的波函数代表一种真实的物理波动，一个个的粒子只不过是这种波动的凝聚的体现。他的看法遭到了以玻尔为代表的很多物理学家的反对。

经典的波动是需要传播介质的，波动事实上就是介质的振动，但是，物质波不需要任何传播介质，因为做验证电子波动性的实验时，是在高真空条件下进行的。因此，很多物理学家都对薛定谔的真实波动图景表示怀疑。

1926年9月，玻尔邀请薛定谔到哥本哈根进行学术演讲，介绍他的新理论。报告结束后，玻尔留薛定谔住了下来，日夜探讨这一理论。讨论过程中，玻尔对薛定谔的波函数诠释发起强烈质疑。我们现在知道，薛定谔的诠释是错误的，可想而知，他面对玻尔的质疑，很难自圆其说。

自己呕心沥血得到的新理论，受到了量子权威人物的质疑，任谁也不能不着急，薛定谔急火攻心，病倒了。他躺在床上，由玻尔夫人照料他的生活。但即便如此，玻尔仍然坐在他的床边，继续对他说："但是你肯定理解，你的物理解释是不充分的……"

薛定谔简直要绝望了，他闭着眼睛，痛苦地说："我真后悔，我为什么要搞这个量子理论……"

玻尔一看情势不对，赶紧安慰道："我们所有人都感谢你。你的波动力学在数学上清晰简单，这是一个巨大的进步。只是，有一些问题是必须搞清楚的……"

那么，如果薛定谔对波函数的物理诠释不正确，到底什么才是正确的诠释呢？

答案很快就揭晓了——概率！

1926年，德国哥廷根大学物理系教授马克斯·玻恩（1882—1970）给出了一个可以让人接受的诠释，他认为，波函数并不像经典波一样代表实在的波动，他只能代表粒子在空间出现的统计规

律：“我们不能肯定粒子在某一时刻一定在什么地方，我们只能给出这个粒子在某时某处出现的概率，因此物质波是概率波，物质波在某一地方的强度与在该处找到粒子的概率成正比。”



玻恩

玻恩给出的波函数的具体的物理诠释是：波函数$\psi(x, y, z, t)$的绝对值的平方$|\psi(x, y, z, t)|^2$代表$t$时刻在空间$(x, y, z)$点发现粒子的概率密度。

在经典物理学中，波的强度正比于振幅的平方。现在，$|\psi|^2$表示概率密度（即概率波的强度），因此波函数$\psi$可以看作是概率波的振幅，简称概率振幅或概率幅。

总结一下，波函数已经出现了几种不同的叫法——波函数、态函数、概率幅；叫波函数是因为它能描述粒子的波动性，叫态函数是因为它能描述粒子的量子状态，而叫概率幅是因为它的平方反映概率波的强度。这三种不同的叫法从不同侧面反映出波函数所蕴含的物理内涵。

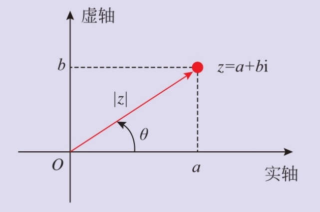
这样，概率作为一种基本法则进入了物理学，物质波只是一种概率波，并非真正的物理波动，波函数只允许计算在某个位置找到某个粒子的概率。对某一物理量进行测量，只能预测出现某一结果的概率，却不能预测一定会得到什么结果。

玻恩找到波函数的概率诠释以后，原子中电子运动的秘密终于被破解了。原子轨道波函数给出的是电子在空间某点的概率幅，波函数的平方决定了电子出现在这个点的概率密度。将波函数的平方作图，就能看出电子在原子核周围空间的概率密度分布，这就是我们通常所说的“电子云”。

电子云的图像并不容易在纸面上表现出来，它本身应该是一个三维空间图像，以原子核为中心，周围空间中每一个点都有一个具体的概率密度数值。为了表现这些数值的大小，人们想到了一个办法，将每一点的概率密度数值与颜色深浅相对应，颜色越深的点表示概率密度越大，越浅的点表示概率密度越小。为了方便观察，通常只画出通过原子核的二维截面，如图5-2所示。如果用一句话来描述核外电子的运动规律，那就是：电子没有固定的运动轨迹，只有概率分布的规律。

## 扩展阅读

波函数在很多情况下都是复数。任意一个复数$z=a+b\mathrm{i}$可以表示成复平面上的一个向量，此向量的长度是$\sqrt{a^2+b^2}$，称为复数$z$的模或绝对值，记为$|z|$（如图5-1所示）。显然，复数的模的平方$|z|^2=a^2+b^2$。如果$z$是一个实数，则$|z|^2=z^2$。




复数在复平面中的表示（$a$是实部，$b$是虚部，i是虚数单位，$\mathrm{i}^2=-1$，$|z|$是模）

值得一提的是，玻恩依靠量子力学，系统地建立了固体的晶格动力学理论。20世纪50年代，玻恩与我国半导体物理奠基人黄昆先生合

著的《晶格动力学理论》，一直是固体物理领域的权威著作。

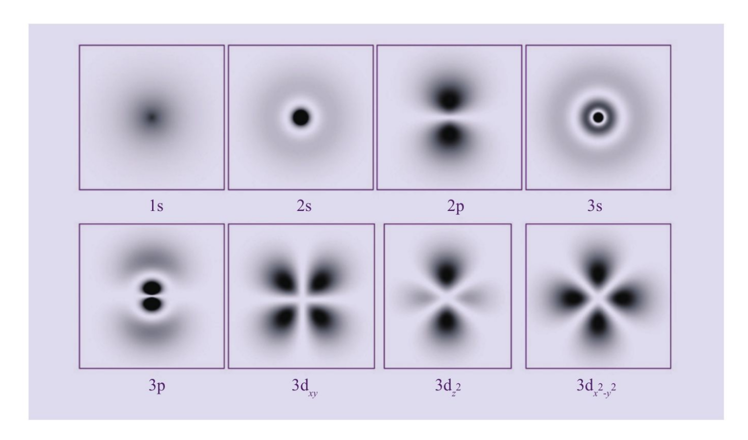



图5-2　不同原子轨道的电子云图



什么是电子云？它是电子在核外出现的概率吗？

---

(1)ψ为希腊字母，读音为/psai/，另一个相似的字母$\varphi$读音为/fai/。

# 6　男孩们的物理学

1925—1928年，是量子力学史上最辉煌的年代，短短几年之间，量子力学理论迎来“井喷式”发展。回顾历史，薛定谔在1926年创建的波动力学，并不是历史上首次出现的量子力学的数学表示形式。事实上，在薛定谔之前大约半年，德国物理学家沃纳·海森伯（1901—1976）已经提出了量子力学的一种数学表示形式，由于它主要依靠矩阵来描述物理量，所以被称为矩阵力学。



海森伯

创建矩阵力学时，海森伯只有24岁，还是一个大男孩。海森伯从小就很有数学天分，13岁时就掌握了微积分。1920年，19岁的海森伯中学毕业，进入慕尼黑大学攻读物理学，师从著名理论物理学家索末菲。在读大学的第一学期，海森伯就对当时还没解决的物理难题——反常塞曼效应提出自己独到的见解，令索末菲刮目相看，直接把他升到研究生班攻读博士学位。海森伯对时间、空间、原子结构、量子理论等大题目很感兴趣，也愿意投入精力研究，但是，索末菲认为海森伯应该加强基础训练，就给他选定了一个流体力学方面的题目作为博士课题进行研究。海森伯对流体力学并不喜欢，但为了毕业，只好硬着头皮搞研究。1923年，他终于写出题为《关于流体流动的稳定

和湍流》的博士论文，虽然博士答辩时磕磕巴巴，有一些简单的问题也没答上来，但总算拿了个及格分勉强过关，取得了博士学位。

博士毕业后，海森伯就抛开流体力学，全身心地投入到量子理论的研究中。事实上，早在博士毕业前一年，他就下定决心要研究量子理论，而这一切，要从他与玻尔的一次散步开始说起。

那是1922年初夏，玻尔应邀到德国哥廷根大学讲学，报告他的原子结构理论。那时候的玻尔，已经是量子学派的掌门级人物，他在丹麦首都哥本哈根建了一个理论物理研究中心，向全世界开放。那是广大青年学子心中的量子圣地。听闻玻尔前来德国演讲，索末菲特意带着他的得意门生海森伯赶到哥廷根去听讲。

海森伯学过玻尔的原子结构理论，知道相关内容，但是听玻尔本人亲自讲，却似乎完全不同了。海森伯清楚地意识到，玻尔所取得的研究成果首先靠的是直觉和灵感，然后才有计算和论证。这让他深受启发。

玻尔的量子理论还存在很多难以解决的困难，在玻尔演讲结束后，海森伯提了一个与玻尔意见相左的问题，这立刻引起了玻尔的注意，发觉这是一个可造之才，演讲结束后，便邀他一起去郊外散步。这次散步，玻尔与海森伯足足谈了3小时，玻尔对海森伯坦诚相见，并不掩饰他对于自己理论的困惑与烦恼，这让海森伯颇为意外，海森伯这才意识到，量子理论才刚刚起步，还有大片的未知领域等待开发，在这一刻，他就下定决心要把发展量子理论作为终生的事业。散步结束后，玻尔邀请海森伯有机会去他那里访问。

这一别，就是两年。1923年，海森伯博士毕业后，先来到哥廷根的玻恩门下担任助手。当时玻恩正在思考如何解决玻尔的理论没法解决的多电子原子的量子化的问题，这正是海森伯感兴趣的方向。

1924年，玻尔给海森伯争取到一笔奖学金，海森伯终于来到玻尔门下，与玻尔一起工作。在玻尔的悉心栽培下，海森伯进步神速。第二年，他就发明了一种用“表格”来处理原子光谱的量子力学方法。当他把论文寄给玻恩看时，玻恩立刻发现，海森伯发明的“表格”其实就是数学中的矩阵，而且海森伯的方法意义重大，据此可以建立一整套

量子力学的新理论。玻恩十分兴奋，他立刻联手另一位数学家约丹，和海森伯一起，很快就发展出矩阵力学理论。海森伯也由此一跃进入顶尖量子物理学家的行列。

## 扩展阅读

矩阵就是一个矩形排列的数值表，一般是$n$行$n$列，例如，下面的$\boldsymbol{A}$和$\boldsymbol{B}$就是两个2行2列的矩阵：

$$\boldsymbol{A}=\begin{pmatrix}1 & 2\\3 & 4\end{pmatrix},\quad \boldsymbol{B}=\begin{pmatrix}5 & 6\\7 & 8\end{pmatrix}$$

矩阵除了可以进行加减乘除这样的计算，它还具有一些特殊的操作，如行列相互调换等。矩阵操作能用于对多个变量在多次观测中的复杂关系进行求解。

矩阵有一个重要的性质就是不满足乘法交换律，例如，上面两个矩阵相乘，$\boldsymbol{AB}\neq\boldsymbol{BA}$（读者可试着找找其中的乘法规则）：

$$\boldsymbol{AB}=\begin{pmatrix}19 & 22\\43 & 50\end{pmatrix}$$
$$\boldsymbol{BA}=\begin{pmatrix}23 & 34\\31 & 46\end{pmatrix}$$

矩阵乘法的不可交换性是量子力学里算符不可交换性的数学基础，会导致完全无法用经典力学理解的量子效应，如海森伯不确定关系（见第7章）。

矩阵力学虽然抢先登场，但是运用的数学太过复杂，物理含义太过抽象，让物理学家们很是头疼。而薛定谔方程是一个偏微分方程，是物理学家们熟悉的数学形式，所以当波动力学出现以后，立刻受到了普遍的欢迎。

短短半年之内，一下子出现两种量子力学，真是让人无所适从，两种看起来完全不同的理论都能解释相同的实验现象，这实在是令人费解的。到底谁对谁错呢？谁的孩子谁心疼，一开始，薛定谔和海森

伯两人都为自己的理论辩护，认为只有自己才是正确的，排斥对方的理论。薛定谔在他的一篇论文中声明："我绝对跟海森伯没有任何继承关系。我自然知道他的理论，但那超常的令我难以接受的数学，以及直观性的缺乏，都使我望而却步，或者说将它排斥。"海森伯也在写给朋友的信中说他发现薛定谔理论是"令人厌恶的"。

但是，要想驳倒对方，就要了解对方。正所谓知己知彼，百战不殆。薛定谔为了驳倒海森伯，开始仔细研究海森伯的理论，结果这一研究才发现，原来是大水冲了龙王庙，一家人不识一家人，这两种理论在数学上竟然是等价的。1926年，薛定谔证明，任何波动力学方程都可变换为一个相应的矩阵力学方程，反之亦然。这一发现终于化干戈为玉帛，此后，两大理论便统称为量子力学。

简单来说，量子力学的这两种数学形式，体现了"波粒二象性"的不同表现：矩阵力学外表描述粒子，将波动性隐藏其中；波动力学则相反，外表描述波动性，而将粒子性隐藏起来。所以二者乍一看好像毫无共同之处，其实是一样的。

1928年，英国物理学家保罗·狄拉克（1902—1984）补上了最后一块拼图，他运用数学变换理论，把波动力学和矩阵力学统一了起来，使其成为一个概念完整、逻辑自洽的理论体系，自此，量子力学终于正式建立起来了。图6-1梳理了量子力学建立的历史脉络。



狄拉克

狄拉克自幼聪颖，16岁就上了大学，读电机工程专业。在大学期间，他对爱因斯坦的相对论非常感兴趣，虽然年纪不大，但是他把广义相对论里的黎曼几何都搞得一清二楚。要知道，黎曼几何是相当艰深的，很多物理学者都望而生畏。而当大多数物理学者还只能欣赏相对论的时候，狄拉克已经在求解引力场方程了。19岁时，狄拉克获得了工程学位，但是，他并没有就业，而是转到了数学系，继续攻读数学学位，很快，数学系的老师和同学就对他的数学能力刮目相看。

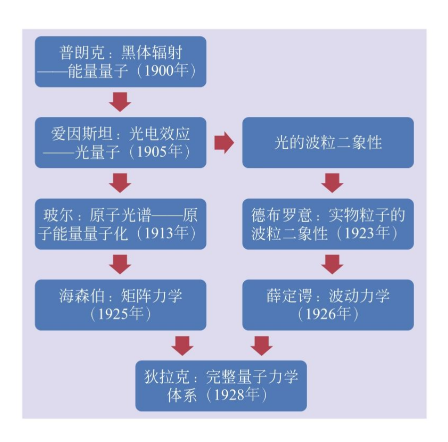



图6-1　量子力学建立的历史脉络

有一次，一位老师正在讲课，黑板上已经写满密密麻麻的符号和公式，同学们都在忙着埋头记笔记。这时候，老师却突然不讲了，盯着黑板发愣，同学们这才发现，课讲不下去了，推导出现了矛盾，没法继续讲了。老师看了好一会儿也没找出原因，只好求助狄拉克："哪里出了错，你能把它指出来吗？"狄拉克不慌不忙地走到讲台上，不但把错误指了出来，还说出如何去更正它。原来，狄拉克早已注意到这个错误。

在数学系的3年，狄拉克不仅打下坚实的数学基础，还从数学的角度反复地对物理学进行了思考。他认识到，要想用最简洁的语言表述自然规律，最好的方法就是利用数学。

1923年，狄拉克到剑桥大学读研究生，研究相对论，第二年毕业后继续留校做研究。1925年，海森伯来到剑桥大学做演讲，他介绍自己最近所写的关于矩阵力学的论文。这次演讲立刻引起了狄拉克的兴

趣，他决定把研究方向从相对论转向量子力学。

1926年9月，狄拉克到玻尔的理论物理研究中心访问，在这里待了半年多。1927年2月，他又到了德国哥廷根大学，在此也待了半年并结识了玻恩等人。同年10月，狄拉克回到剑桥。这时候，恐怕没有人意识到，这位年仅25岁的年轻人，已经对相对论和量子力学了如指掌。正所谓厚积薄发，第二年，他就做出了诺奖级别的贡献，他把相对论引入量子力学，建立了狄拉克方程。

狄拉克早就发现，薛定谔方程首先不具备相对论条件下的协变性质，不适用于超高速运动的粒子；其次没有把电子的自旋性质囊括进去，而电子的自旋，就像它的质量、电量一样，也是电子的重要性质。经过不断地尝试，他终于发现了相对论形式的薛定谔方程，也就是狄拉克方程。这一方程不但解决了上述两个问题，还预言了反物质的存在，使量子力学理论登上了一个新的高度。

霍金曾这样评价狄拉克的贡献："狄拉克阐述了任何系统的量子力学的一般规则，这些规则结合了海森伯和薛定谔的理论并指出它们的等价性。在现行量子力学的三个奠基人中，海森伯和薛定谔的功劳使他们各自看到了量子理论的曙光，但是正是狄拉克把他们看到的交织在一起，并揭示了整个理论的图像。"

## 扩展阅读

### 与牛顿方程比肩——薛定谔方程

薛定谔方程在量子力学中的作用，相当于牛顿方程在经典力学中的作用。处理量子力学问题，首先就是写出薛定谔方程，然后进行求解，可解出能量与波函数，进而可求其他可观测量。正是通过对薛定谔方程的求解，人们认识到了微观世界许多奇异的量子特性。所以，让我们一起来欣赏一下这个伟大的方程。

薛定谔方程是由两个能量算符作用于波函数上构成的恒等式：

$$\hat{H}\psi = \hat{E}\psi$$

式中，$\hat{H}$和$\hat{E}$都是能量算符；$\hat{H}$是用动能与势能之和表示的能量算符，也叫哈密顿算符；$\hat{E}$是用时间表示的能量算符；ψ是体系的波函数，即ψ（$x, y, z, t$）。如果给定粒子的初始状态，就可以通过薛定谔方程求解出任一时刻的状态，也就是说，薛定谔方程描述了波函数随时间的演化过程。

上面的式子看起来很简单，但是如果把两个能量算符的具体形式代入，它就变成了下面的样子：

$$\left[-\frac{\hbar}{2m}\left(\frac{\partial^2}{\partial x^2}+\frac{\partial^2}{\partial y^2}+\frac{\partial^2}{\partial z^2}\right)+\hat{V}\right]\psi = \mathrm{i}\hbar\frac{\partial}{\partial t}\psi$$

式中，$\hat{V}$是势能算符；$\hbar=h/2\pi$，称为约化普朗克常数（因为量子力学中经常用到$h/2\pi$这个数，为了书写方便，将其记为$\hbar$）。

这是一个复杂的偏微分方程，对于本书的读者来讲，没有必要去深究其数学上的细节，我们只要知道这是一个微分方程就行了。微分方程是牛顿的天才性创造，它可以把一个复杂的运动过程分解为无穷多个微小的部分来研究，微分方程也因此成为物理学中最基本的方程形式。无论是牛顿力学，还是量子力学和相对论，都离不开微分方程。

看到这儿，读者可能心里还在疑问，到底什么是算符呢？算符其实并不难理解，它就是把一个函数变成另一个函数的运算符号，如相乘、开方、求导等都是算符。根据量子力学的算符假设，微观体系的每一个“可观测量”（如坐标、动量、角动量、能量等）都与一个算符相对应，算符用该物理量加一个倒三角来表示（例如，坐标$x$的算符记为$\hat{x}$，动量$p$的算符记为$\hat{p}$，势能$V$的算符记为$\hat{V}$）。

读者要问了，为什么要规定这些算符呢？在此处，引入算符的目的是运算。算符的运算对象主要是波函数，根据基本假设，把一个物理量的算符作用在波函数上进行运算，如果结果正好等于一个常数乘以这个波函数，那么这个常数就是这个物理量的本征值，这个波函数就叫做本征态。

例如，把能量算符（哈密顿算符）作用在氢原子的1　s轨道波

函数上，正好等于-13.6 eV乘以1 s轨道波函数，即

$$\hat{H}\ \psi_{1s}=-13.6\ \text{eV}\times\psi_{1s}$$

那我们就说，氢原子1 s轨道电子的能量等于-13.6 eV，这是能量的本征值，$\psi_{1s}$是能量的本征态。

每一个算符都对应一系列本征态和本征值，本征值对应着该物理量的可能的观测结果。在本征态下测量此物理量，将测得确定的本征值；在非本征态下测量此物理量，测量结果不确定，但必为某一个本征值，且测量以后波函数坍缩到该本征值对应的本征态。

在量子体系的诸多状态之中，有一类特殊的状态，那就是能量取确定值的状态，称之为定态。定态下能量的取值不随时间变化，概率分布也不随时间变化，对应的波函数$\psi(x, y, z)$不含时间，称为定态波函数。于是含时薛定谔方程式可退化为定态薛定谔方程：

$$\hat{H}\psi = E\psi$$

式中，$E$为体系能量。可以看到，定态薛定谔方程就是哈密顿算符满足的本征方程，求解此方程，即可得到体系的能量本征值与本征态波函数，从而了解体系的量子力学运动规律（想进一步深入了解的读者可参阅附录）。

不难看出，哈密顿算符是薛定谔方程的根基，神奇的是，哈密顿量（体系的动能和势能总和）也是经典力学的根基。哈密顿量是由爱尔兰数学家哈密顿提出来的，他从哈密顿量出发严格地推导出了牛顿力学，从而把牛顿力学纳入一个新的数学框架中。量子力学颠覆了牛顿力学，但是哈密顿量不但没有被量子力学抛弃，反而提升为哈密顿算符，成为量子力学的基础。

波函数、薛定谔方程、算符、本征值与本征态，这些都是量子力学的基本假设，类似于几何学中的公理，没法证明，但是由此推出的所有结论都能很好地解释和预测实验结果，所以得到了大家的公认。



量子力学的五个基本假设

# 第三篇　量子·颠覆认知

# 7　无迹可寻

虽然薛定谔是波动力学的创始人，但波函数的解释权已经完全脱离了薛定谔预定的轨道。人们普遍接受了玻恩对波函数的概率诠释。波函数只允许计算在某个位置找到某个粒子的概率，对于体系的演化，只能预测某一结果的概率，却不能预测一定会得到什么结果。概率作为一种基本法则进入了量子力学。

但是，包括薛定谔在内，还有许多物理学家对量子力学这种固有的不可预测性持怀疑态度，其中的带头人便是爱因斯坦。在经典力学中你掷一个骰子（图7-1），我们说你只能预测一个概率，每个面朝上的概率都是1/6，但这是因为我们忽略了很多物理细节，例如，抛骰子的力度、角度、手法，骰子自身的弹性、棱角、密度分布，以及空气分子的分布情况、桌面材料的弹性，等等，如果你把所有因素都考虑进去，所有细节都能掌握的话，那么骰子抛出的一瞬间，就可以准确地预测它的运动轨迹，因此也一定能准确地预测它到底哪面朝上，这时候，结果是确定的，不存在概率性。因此，持怀疑态度的物理学家认为量子力学中的概率性也是因为测量不精的原因，如果你能把所有细节都测量出来的话，它就不会呈现概率性。这种想法似乎有一定道理，但是，海森伯在1927年发现了一条量子力学的基本原理，直接否定了这种想法。



图7-1　经典的概率统计——掷骰子

海森伯发现的原理叫不确定原理：有一些成对的物理量，要同时测定它们的任意精确值是不可能的，其中一个量被测得越精确，其共轭量就变得越不确定。例如，坐标与相应的动量分量、能量与时间等（两个共轭量相乘后的单位正好是普朗克常数的单位J·s）。

对于坐标与相应的动量分量，不确定原理的数学表达式是

$$\Delta x \cdot \Delta p_x \geqslant \frac{\hbar}{2}$$

上面的关系式表明，在量子力学里，一个粒子不可能同时具有确定的位置和速度，一个粒子的位置测得越精确，它的速度就越不精确，反之亦然。因此，在测量粒子的位置和动量时，它们的精度始终存在着一个不可逾越的限制，也就是说，你不可能准确地测量一个粒子的运动轨迹，这不是测量仪器的精度问题，而是自然界的根本属性，这样的话，它的概率性就成了必然。因此，量子力学里的概率和经典力学里的概率是不一样的，经典力学里的概率来自于测量的不精确，而量子力学里的概率来自于体系自身的内在属性。或者说，量子力学的概率内涵是绝对的。

从某种意义上来说，正是量子力学的概率内涵，让我们的人生充满着不确定性。作为由天文数字的基本粒子组成的集合体，人体内部每个粒子的即时运动都是不可预测的，因此我们的未来也是不可预测的，这样，我们自身的努力才是有意义的。如果一切都是决定性的，那我们的人生就像演电影一样，只能按照已经预定好的剧本一路演下去，那人生的意义何在呢？所以说量子力学的不确定性对人类来说是幸运的，它让我们避免成为大自然的提线木偶，而让我们成为自身命运的主宰。

因为物理学界对量子力学的概率诠释一直存在争议，所以量子力学自诞生以来经历了各种非常严格的实验检验，但到目前为止，还没有发现任何能够推翻量子力学的实验证据。

# 8　不一样的骰子

虽然我们经常用掷骰子来作为量子力学概率性的比喻，但是，如果你掷的不是一个骰子，而是多个骰子的话，你可能不会想到，量子力学中的概率统计和经典物理中的概率统计是不一样的。而这一发现，来自于一次“错误”的授课。

1922年，印度达卡大学物理系讲师纳特·玻色（1894—1974）正在给学生讲授黑体辐射，讲授过程中，他以爱因斯坦提出的光量子为对象，运用经典的麦克斯韦-玻尔兹曼统计来推导公式，打算向同学们展示公式推导过程中的疑难。但是，当他推导完毕以后，结果让他目瞪口呆，他竟然推导出了普朗克的黑体辐射公式！

要知道，普朗克当年凑出黑体辐射公式以后，虽然自己给出了一个推导并首次提出量子化的概念，但他的推导是存在严重缺陷的。后来，爱因斯坦给出了一个新的推导，但也并不是完美无缺的，也就是说，那时候还没有人能从理论上完美地推导出普朗克公式。现在，玻色竟然无意间完成了这一重大发现。

昨天备课时候还不是这样的，今天怎么变了呢？玻色仔细检查了他的推导过程，这才发现，他在讲课时，由于疏忽，犯了一个“小错误”。

我们可以举个最简单的例子来展示他犯的错误。假设加热黑体的时候，黑体会辐射出两个光子，每个光子都有50%的概率处于频率$\upsilon_1$，50%的概率处于频率$\upsilon_2$，那么请问，两个光子都处于频率$\upsilon_1$的概率有多大？

这个问题如果用经典概率统计来计算，很简单，答案为25%。因为有以下四种情况如图8-1（a）所示。

| 第 1 种可能 | 两个光子都处于 $\upsilon_1$ |
|---|---|
| 第 2 种可能 | 两个光子都处于 $\upsilon_2$ |
| 第 3 种可能 | 光子 1 处于 $\upsilon_1$，光子 2 处于 $\upsilon_2$ |
| 第 4 种可能 | 光子 1 处于 $\upsilon_2$，光子 2 处于 $\upsilon_1$ |

但是，玻色在黑板上推导的时候，却犯了一个“错误”，他只考虑到了以下三种情况如图8-1（b）所示。

| 第 1 种可能 | 两个光子都处于 $\upsilon_1$ |
|---|---|
| 第 2 种可能 | 两个光子都处于 $\upsilon_2$ |
| 第 3 种可能 | 两个光子一个处于 $\upsilon_1$，另一个处于 $\upsilon_2$ |

这样，按照他“错误”的推导，两个光子都处于$\upsilon_1$的概率变成了1/3。这样误打误撞，竟然推导出了黑体辐射公式。

找到了自己的“错误”，玻色立刻意识到，这里面大有玄机！经过仔细分析，他发现这两种概率的区别就在于：经典统计里，两个光子是可以区分的，而在自己新的统计中，两个光子是不可区分的！

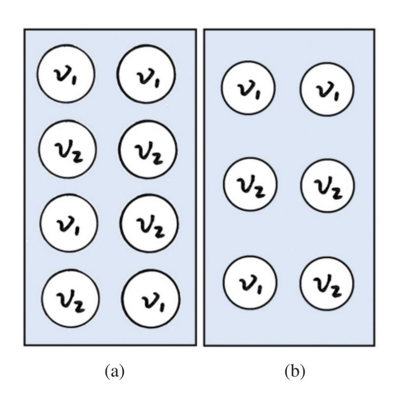



图8-1　经典统计和玻色统计的区别

（a）经典统计；（b）玻色统计

这让他大为兴奋，把自己的发现写了一篇论文，题目是《普朗克定律与光量子假说》。在文中，玻色指出经典的麦克斯韦-波尔兹曼统计不适合于微观粒子，并用自己提出的所有光子不可区分的假设，推导出了普朗克定律。

然而，当他把论文投递给英国一家知名杂志后，主编认为玻色犯了十分低级的错误，论文毫无价值，直接退稿了。无奈之下，玻色想到了爱因斯坦，他决定直接将论文寄给爱因斯坦，向爱因斯坦求助。

1924年6月，爱因斯坦收到了玻色的论文。爱因斯坦立刻意识到玻色的发现具有重大意义。他亲自将玻色的论文翻译成德文，并将其推荐给德国最主要的物理刊物。同时，受玻色工作的启发，爱因斯坦自己也写了一篇关于光子统计的论文，两篇文章在同一刊物一起发表出来。这种新的统计方法后来被人们称为玻色-爱因斯坦统计。

在量子统计中，由于相同的粒子具有不可区分性，因此被人们称为全同粒子。在经典物理学中，是没有全同粒子的，因为经典物理学中的粒子都具有明确的运动轨迹，是可以明确区分的。而量子理论中，由于不确定原理的限制，粒子没有明确的运动轨迹，再加上粒子波函数在空间中的重叠，导致各个粒子完全没法区分。例如，氦原子里有两个电子，这两个电子就是全同粒子，假如将两个电子交换位置，这个氦原子看不出任何状态上的变化。

不久以后，人们又发现，即使是全同粒子，它们的统计规律也不同，例如，电子就和光子不同。举例来说，电子有个性质叫自旋，既可以自旋向上，也可以自旋向下。对于基态氦原子，它的两个电子都处于1s轨道上，如果电子的统计规律像光子一样，那么这两个电子的自旋状态就有三种可能：两个都向上、两个都向下、两个正相反（图8-2（a））。但事实上，这两个电子的自旋状态只有一种可能：两个正相反（图8-2（b））。

电子的这种统计规律叫费米-狄拉克统计。为什么会这样呢？原因就在于，电子还受另一个原理的制约——泡利不相容原理。

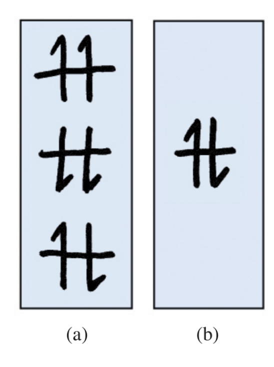
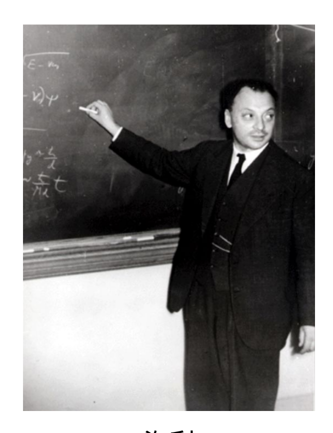



图8-2　玻色统计和费米统计的区别

（a）玻色统计；（b）费米统计



泡利

沃尔夫冈·泡利（1900—1958）是海森伯的师兄，他也师从慕尼黑大学的索末菲。1918年，18岁的泡利中学毕业，他不想读大学，觉得太浪费时间，就直接去找索末菲，要读他的研究生。不可思议的是，索末菲对他进行面试以后，居然同意了，这样，一个中学生直接成了研究生。

但是，索末菲可不是胡闹，泡利是真的有才华。1921年，德国的《数学科学百科全书》邀请索末菲撰写关于相对论的一卷，索末菲无暇撰写，就推荐了泡利，编委会出于对索末菲的信任，就同意了。结果，21岁的泡利很快就写好了一篇200多页的相对论介绍，精辟地论述了狭义和广义相对论的数学基础和物理原理，很多地方都有自己独

到的见解。

书出版后，索末菲给爱因斯坦寄了一本，爱因斯坦读后，大加赞赏，当他得知作者仅仅21岁，还是一个学生，更是吃惊。他评价道：“任何该领域的专家都不会相信，该文竟然出自一个年仅21岁的年轻人之手。作者对这个领域的理解力、熟练的数学推演力、深刻的物理洞察力、表述问题的清晰性、系统处理的完整性、语言把握的准确性，会使任何一个人都感到羡慕。”

1921年，泡利博士毕业，被索末菲推荐到哥廷根大学的玻恩门下做助手。1922年，玻尔到哥廷根大学讲学，在和泡利接触后，很欣赏他的才华，就邀请泡利到哥本哈根访问。于是泡利又到哥本哈根在玻尔门下工作了一年。读者还记得，那一次，玻尔还向海森伯发出了邀请。玻尔相中的这两个年轻人，后来都成为哥本哈根学派的领军人物。很有意思的是，海森伯的求学之路几乎就是跟在泡利后面步步紧随，他比泡利晚一年师从索末菲，也比泡利晚一年给玻恩当助手，还比泡利晚一年去哥本哈根访问。最令人称奇的是，他们俩都是中学毕业3年后就拿到了博士学位。

玻尔曾经提出一个问题——如果原子中电子的能量是量子化的，这些电子为什么没有都排布在能量最低的轨道呢？如果你观察元素周期表，就会发现每一种元素原子的电子排布都不相同，随着电子数的增多，电子排满了从低到高的各个能级。玻尔对此很不解，因为自然界的普遍规律是一个体系的能量越低越稳定（这叫能量最低原理），这些电子为什么要往高能级排布呢？

这个问题最终被泡利所解决。1925年，泡利根据对原子经验数据的分析提出一条原理：原子中任意两个电子不可能处于完全相同的量子态。这就是泡利不相容原理。

泡利不相容原理是一个非常重要的理论，正因为如此，电子才会乖乖地从低能级到高能级一个一个往上排列。也正因为如此，电子才会构成一个个不同的原子，从而出现我们看到的五彩缤纷的元素。

人们发现，微观粒子有的受泡利不相容原理的制约，有的不受。因此，微观粒子的统计规律分为两种：一种是像光子那样不受泡利不

相容原理的制约的粒子，满足玻色-爱因斯坦统计；另一种是像电子那样受泡利不相容原理的制约的粒子，满足费米-狄拉克统计。如前所述，玻色-爱因斯坦统计是在1924年提出来的，而费米-狄拉克统计是在1926年由狄拉克和意大利物理学家费米各自独立地提出来的。

现在，全同粒子已经作为一条基本假设被纳入量子力学的理论框架，人们发现的各种实验现象都证明了该假设的正确性。

# 9 没有人能理解

2002年，美国两位学者在美国的物理学家中做了一次调查，请他们提名史上十大最美物理实验。最终，电子双缝干涉实验排名榜首。这个实验为什么受到如此青睐呢？原因就在于，这个实验展现了概率波谜一般的特征。用量子力学大师费曼的话说，就是“量子力学的一切，都可以从这个简单实验的思考中得到”。正因为如此，费曼在他所著的《费曼物理学讲义》中，把电子双缝干涉实验放在量子力学的开篇进行讲述，可见这个实验对于理解量子力学的重要性。

在经典物理学中，波是比较好理解的概念。多个波可以在同一空间中同时存在，并且发生叠加，产生干涉现象。干涉条纹是波与波的叠加产生的波动加强或抵消的结果：波峰和波峰叠加，波动加强，波峰和波谷叠加，波动抵消。图9-1给出了常见的正弦波的波形，以及它们进行干涉叠加时加强或抵消的图像。

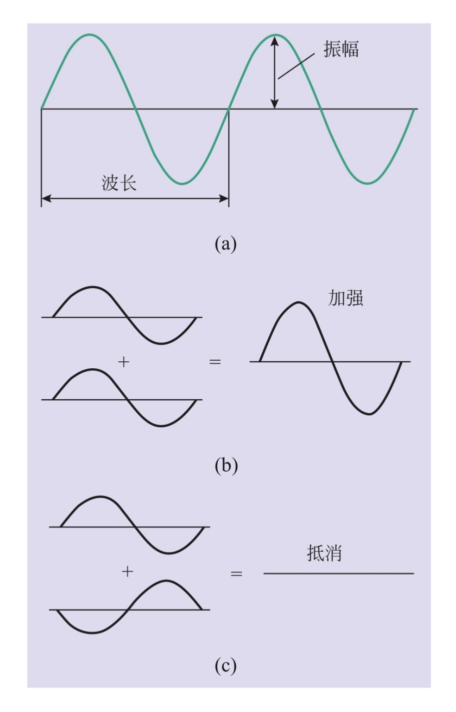



图9-1　正弦波以及其干涉叠加示意图

（a）正弦波波形；

（b）相长干涉；

（c）相消干涉

可以说，干涉是波动最重要的特征，而干涉现象最典型的例子就是双缝干涉实验。以常见的水波为例，在一个水槽中用一个上下振动的小球作为波源，在水面产生圆形的波，在这个波前方放一块木板，木板上刻有两条狭缝，入射波在狭缝处发生衍射，形成两列新的圆形波，这两列波就会发生干涉。如果在后面放一块探测屏来测量干涉波的强度，就会显示出明暗相间的条纹，如图9-2所示。

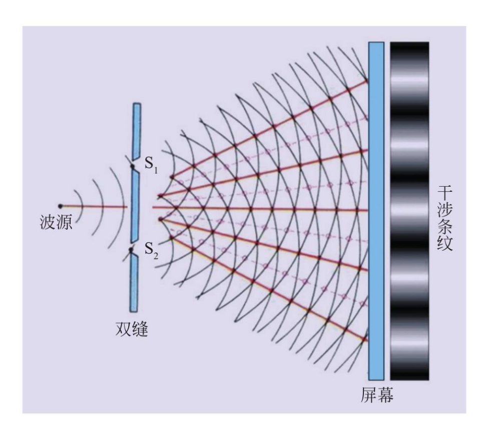



图9-2　水波的双缝干涉实验

18世纪，科学家们为光究竟是粒子还是波争得不可开交，莫衷一是。1807年，英国科学家托马斯·杨（1773—1829）做了一个轰动一时的实验——杨氏双缝干涉实验。他把一束单色光照射到两条平行狭缝上，结果在两条狭缝后面的屏幕上出现了明暗相间的条纹（图9-3），这不就是波的干涉条纹吗？对干涉波的强度进行测量，发现其变化规律与水波完全一致，这就证明光的确是波。

但是，经典的波动是需要传播介质的，人们绞尽脑汁，才为光找到了一种传播介质——“以太”。“以太”最早由古希腊的亚里士多德提出，他设想“以太”是充满天地间的一种媒质，这完全是一种凭空假想，没有任何根据，但光学家们却把它拿来当作光的传播介质。

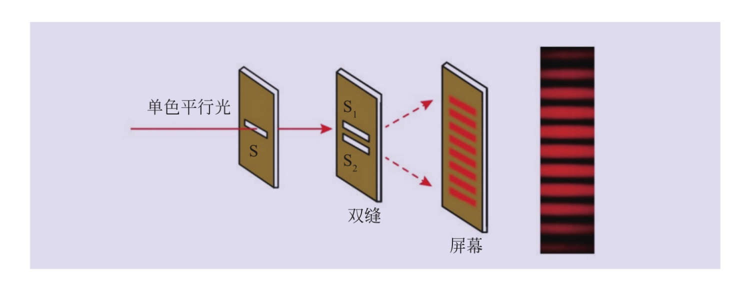



图9-3　杨氏双缝干涉实验示意图

为了寻找“以太”，人们又费尽了心机，但到了1905年，爱因斯坦在相对论中却直接否定了“以太”的存在。这就表明光可以直接在真空

中传播，不需要任何传播介质，和普通的波不一样。

不需要传播介质，这是光波与经典波的最大区别。经典波的能量传递靠介质振动，现在没有介质了，意味着光波的能量只能由它自己携带，所以光是具有粒子性的，光波的能量由一个个光子携带。在经典波中，波的强度取决于介质的波动幅度，而光的强度则取决于光子流的密度。由此看来，正因为爱因斯坦坚决否认“以太”的存在，所以光子的概念由他首先提出来，也就是顺理成章的事情了。

不需要传播介质，这就意味着光波和水波的干涉原理并不一样。对于水波来说，如果水波的波动逐渐减弱，那么干涉条纹也会相应减弱。但不管波动如何微弱，整个水面都在上下振动，水波总是充满整个水面，干涉条纹也布满整个屏幕，只不过是比原来微弱罢了。但是，如果是光波逐渐减弱，会出现什么结果呢？

1909年，英国科学家泰勒做了一个实验。他先用强光照射缝衣针，拍下针孔的衍射图像，再把光源衰减到极弱，结果发现，短时间内并不能出现衍射图像，只有散乱的光点。但是，当他把实验时间延长到2000 h以后，衍射图像又出现了，而且和用强光源得到的图像完全一样。这个实验可以称得上是单光子衍射实验。后来，人们又做出了单光子的双缝干涉实验：光源一次只能发出一个光子，在屏幕上也只能出现一个落点，但是，随着一个个落点的出现，干涉条纹竟然逐渐显现出来。这两个实验都表明，光的波动性是由光子的概率分布体现出来的，现在我们知道，这就是量子力学的概率波。

从光的双缝干涉实验形成的干涉条纹来看，概率波和经典波的干涉条纹强度分布规律是一样的，也就是说，在很多情况下，如果把光看成是弥散在空中的波也没有什么问题，这时候你可以不去考虑光子，而把它看作是波（电磁波理论就把光看作是电场与磁场的振动），这就是绝大多数光学问题都可以用经典波动理论解释的原因。但是，对于少数问题，经典波动理论没法解释，如前面提到的光电效应，这时候，你就得把光看作是光子流了。正是从这个意义上来说，光的行为既像是波，又像是粒子，于是，人们只好给它这种奇怪的性质起了这样一个奇怪的名字——波粒二象性。事实上，如果我们叫它“波粒二不像”，也没什么问题，因为从整体来看，它的性质既不像

波，也不像粒子。

1927年，当通过电子的衍射实验（见第3章）证明了物质粒子也有波粒二象性之后，人们就希望实现电子的双缝干涉实验。但是对于电子来说，由于其波长很短，所以需要很窄的狭缝才行，而要将狭缝做得非常精细是很困难的，这就导致这个实验很难做。

直到1961年，才由德国的约恩逊成功完成了这个实验。他在铜箔上刻出长50 μm，宽0.3 μm，间距1 μm的狭缝，采用50 kV的加速电压，在高真空环境下，使电子束通过双缝，得到了干涉图样。对干涉条纹的强度分布进行测量，发现它和光波的规律是一致的。事实上，电子波和光波的物理本质也是一样的，都属于概率波。后来，人们又成功做了中子、原子和分子的双缝干涉实验，证明了波粒二象性的普遍性。

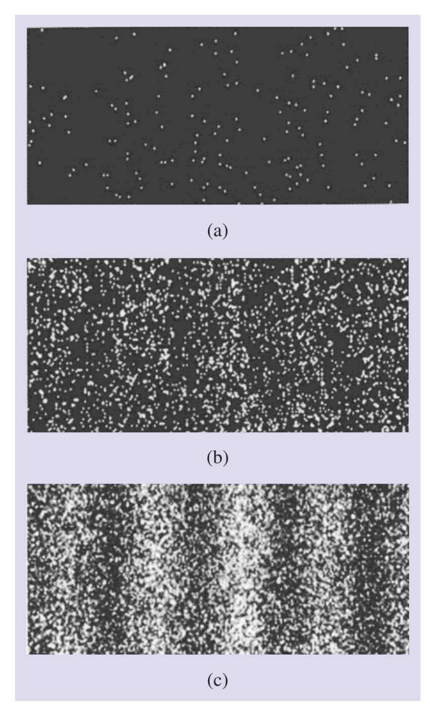



图9-4　单电子双缝干涉实验的细节

再后来，单电子双缝干涉实验也成功了。这是把电子的发射速度

调慢，慢到一次只发射一个电子，等前一个电子落在屏幕上再发射下一个电子，从而保证电子之间相互没有影响，实验结果如图9-4所示。从图中可以看出，刚开始，每个电子的落点都是随机的，但是不久你就会看出规律，因为屏幕上居然慢慢地出现了干涉条纹，最后，明暗相间的干涉条纹越来越清晰地显现出来。

这就是电子双缝干涉实验排名十大最美实验榜首的原因，因为所有电子之间相互都没有联系，但它们最后一个个重叠起来就形成了干涉条纹！

这个实验结果意味着，每个电子事实上都在按照波动的特征运动，它自己跟自己发生干涉，所以它会落在干涉条纹的位置。也就是说，单个粒子也能表现出波动性，波粒二象性是一种整体性质。

物理学家是刨根问底的一群人，他们接下来迫切地想搞明白一个问题——电子到底是从哪条狭缝穿过去的？按我们通常的想法，即使双缝同时打开，电子的运动也只有两种可能：通过狭缝1，或者通过狭缝2。但是如果是这样，那就应该是两个单缝衍射图像的叠加，而不是得到干涉条纹。要想获知真相，除了“看一看”，似乎别无他法，但是，你要想看它，就得设计一个观察装置。在《费曼物理学讲义》中，费曼提出了这样一个观察装置（图9-5），紧贴双缝后面放一个光源，光源会持续发出光子，当有电子从旁边经过时，被它散射的光子会被光子探测器捕捉到，从而可以断定电子从哪条缝通过。假如电子从缝1穿过，会探测到缝1附近有闪光；假如电子从缝2穿过，则会探测到缝2附近有闪光。

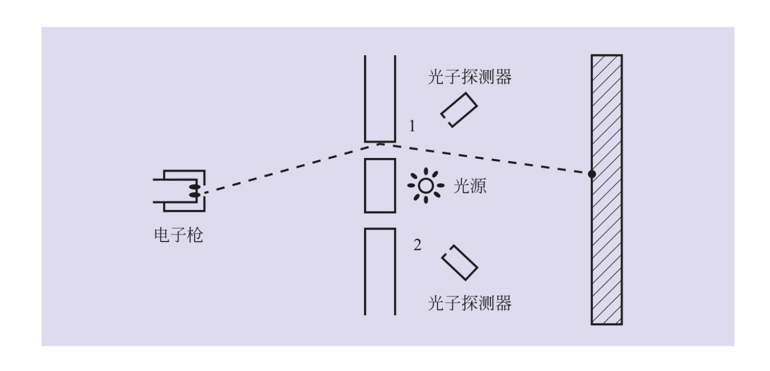


图9-5　观察电子通过哪条缝的实验示意图

如果你去做这个实验，会是什么结果呢？你会看到，或者缝1处有闪光，或者探测到缝2处有闪光，你能判断电子从哪条狭缝穿过。但是，别高兴得太早，你会发现，此时屏幕上的干涉条纹竟然消失了！你看到的只是两个单缝衍射图案的叠加！也就是说，如果我们观察到了电子的路径，它就不再干涉；而如果我们不观察，它就保持干涉。电子好像在跟我们玩捉迷藏的游戏，就是不让你知道它是如何自己跟自己干涉的，只能让人徒唤奈何！

在这个实验中，我们没法确定电子的运动轨迹，唯一合理的解释就是：它没有运动轨迹！粒子没有固定的运动轨迹，只有概率分布的规律，这是量子力学中粒子运动的普遍规律。事实上，这也是不确定关系的必然结果，如果有轨迹，动量和位置就同时确定了，就不满足不确定原理了。

这个实验太过不可思议，所以直到现在，还有物理学家在用不同的手段研究这个实验，希望从中破解量子的奥秘。



不可思议的双缝干涉实验，没有人能理解量子力学！

# 10 上帝掷骰子吗

1927年10月，第五次索尔维会议在比利时布鲁塞尔召开，会议主题为“电子和光子”。这次会议的与会者29人中，有17位诺贝尔奖得主，量子理论的创始人几乎全数出席，可谓是物理学史上绝无仅有的巅峰阵容（图10-1）。这时候，波动力学和矩阵力学已经诞生，概率诠释和不确定原理也被提出，量子力学似乎一夜之间就形成了一套完整的理论体系，但是，其对经典物理观念的冲击太大，以至于这些科学巨人们也难以形成统一的意见。因此，在这次会议上，以玻尔为首的概率论支持者和以爱因斯坦为首的决定论支持者展开了激烈的论战，成为科学史上的一段佳话。

这次会议有五个重磅报告：德布罗意的导波理论、玻恩的波函数概率诠释、海森伯的不确定原理、薛定谔的波动方程，以及玻尔的关于量子力学诠释的总结报告。

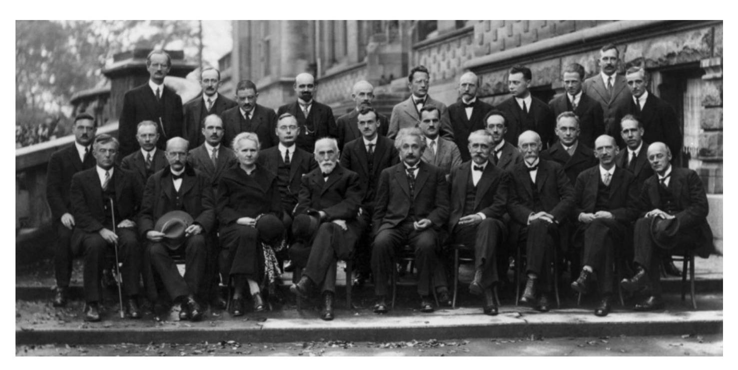


图10-1　第五次索尔维会议合影

第三排：（1）奥古斯特·皮卡尔德；（2）亨里奥特；（3）保罗·埃伦费斯特；（4）爱德华·赫尔岑；（5）西奥费·顿德尔；（6）埃尔温·薛定谔；（7）维夏菲尔特；（8）沃尔夫冈·泡利；（9）维尔纳·海森伯；（10）拉尔夫·福勒；（11）莱昂·布里渊

第二排：（1）彼得·德拜；（2）马丁·努森；（3）威廉·劳伦斯·布拉格；（4）亨德里克·克雷默；（5）保罗·狄拉克；（6）阿瑟·康普顿；（7）路易·德布罗意；（8）马克斯·玻恩；（9）尼尔斯·玻尔

第一排：（1）欧文·朗缪尔；（2）马克斯·普朗克；（3）玛丽·居里；（4）亨德里克·

洛伦兹；（5）阿尔伯特·爱因斯坦；（6）保罗·朗之万；（7）查尔斯·古耶；（8）查尔斯·威尔逊；（9）欧文·理查森

爱因斯坦会前也收到了邀请他作报告的信函，但他拒绝了。他在回信中说："现在这件事我尚不能胜任。在量子理论的近期发展中，我还不具备足够的才智，还不能跟上这狂风暴雨般的进展。此外，另一个原因是，我还不赞成这个以纯统计性为基础的新理论的思考方式。"

从爱因斯坦的回信就能看出他的态度，显然，他对量子力学的概率内涵是很抵触的。爱因斯坦是物理决定论的支持者，他希望人类能对世界给出一个明确的解释，而不是像概率的、不确定的之类在他看来含糊不清的字眼。和他持相同观点的还有德布罗意和薛定谔，他们都反对量子力学的概率诠释。

在爱因斯坦看来，量子力学的概率诠释只是一个权宜之计，那是因为人们还没有能力认识量子背后更深层次的世界的本质，而不是说世界本来就是这样的。而持概率论的玻尔等人则认为世界本来就是概率性的和不确定的，这就是世界的本质。和玻尔持相同观点的有玻恩、海森伯、狄拉克、泡利等人，因为他们都和玻尔交往密切、渊源颇深，所以被称作哥本哈根学派。

虽然爱因斯坦持决定论的态度，但是他忙于相对论，并没有提出相关理论，反倒是德布罗意搞了一个新理论。在会上，德布罗意宣读了论文《量子的新动力学》，提出了一个替代波函数概率诠释的新方法，这个方法他称之为"导波理论"。在导波理论中，德布罗意认为，粒子和波是同时存在着的，粒子就像冲浪运动员一样，乘波而来，在波的导航下，粒子从一个位置到另一个位置，它是有路径的。但是德布罗意刚讲完，导波理论就遭到了泡利的猛烈抨击。泡利是出了名的"毒舌"，习惯于挑剔，善于发现别人演讲中的漏洞，批评起来丝毫不留情面。他有个外号叫"上帝之鞭"，可见同行们多么忌惮他的批评。据说，有一次爱因斯坦做一个关于相对论的讲座，泡利坐在最后一排，当爱因斯坦讲完以后，泡利站起来，直接提了一个尖锐的问题，让爱因斯坦都难以回答。从此以后爱因斯坦每次做演讲都要习惯性地往后排扫一眼，看看泡利在不在场，真是"一朝被蛇咬，十年怕

井绳”。

泡利善于发现问题，目光犀利，导波理论的缺点马上就被他发现了。泡利当场指出，这个理论虽然能解释双缝干涉实验，但在考虑两个粒子碰撞散射时，理论就会瞬间崩溃，更遑论复杂的多粒子系统。面对泡利连珠炮般的攻击，德布罗意左支右绌，难以招架，很快败下阵来，只好将求助的目光投向爱因斯坦，希望爱因斯坦能帮他说几句话。爱因斯坦虽然内心支持德布罗意，但是泡利的指责的确在理，自己也没有好的办法辩解，他也看出来导波理论的确存在明显漏洞，无法使人信服，只好沉默不语。

德布罗意失望地走下讲台，接下来，整个会场就成了哥本哈根学派表演的舞台，量子力学大放异彩。最后玻尔的总结报告做完以后，原本不打算发言的爱因斯坦实在坐不住了，决定发起反击。

他站起来说：“很抱歉，我没有深入研究过量子力学，不过，我还是愿意谈一谈一般性的看法。”然后，他开始对玻恩的概率诠释发难，指出在双缝干涉实验中，如果按照概率诠释，电子落点概率将分布在一个很大的范围内，但是一旦电子落在屏幕上某一点，这一点概率瞬间突变为1，与此同时，其他所有点的概率将瞬间突变为0，那么，这个因果关系的变化速度是超光速的，违反了相对论中的光速极值原理。

面对爱因斯坦的这一指责，玻尔的回应是，瞬间变化的是波函数，而波函数并不是一个真正在三维空间中运动的波，因此不受定域性的束缚。

爱因斯坦没有过多纠缠，会议进行了简单的讨论后就结束了。但是，爱因斯坦并没有罢休，他构思了一夜，在第二天吃早餐的时候，又对不确定原理展开了质疑。他向玻尔抛出了一个思想实验，指出在双缝干涉实验中，如果把双缝吊在弹簧上，就可以通过弹簧测量粒子穿过双缝时的反冲力，从而确定粒子到底通过了哪条狭缝。

玻尔吃了一惊，他花了一整天的时间考虑，到晚餐时，他终于指出了爱因斯坦推理中的缺陷：爱因斯坦的演示要管用，就必须同时知道两个狭缝的初始位置及其动量，而不确定原理限定了同时精确测定

物体的位置和动量的可能性。通过简单的运算，玻尔能够证明，这种不确定性将大到足以使爱因斯坦的演示实验失败。

这一次过招，玻尔胜了。但是，他并没有说服爱因斯坦。爱因斯坦不是一个能轻易被别人左右的人，他只相信自己的物理直觉。

就是在这次会议上，爱因斯坦当众抛出了那句名言："我相信，上帝是不会掷骰子的。"玻尔的回答是："爱因斯坦，不要告诉上帝应该怎么做。"

三年时光转瞬即过，1930年10月，第六届索尔维会议继续召开，这次的主题是"关于物质的磁性"。但这次会议被世人牢记的并不是"磁性"，而是爱因斯坦和玻尔的第二次论战。这一次，爱因斯坦有备而来，主动向不确定原理再次发起挑战。不同的是，他这次没有攻击坐标-动量的不确定关系，而是换成了时间-能量不确定关系。

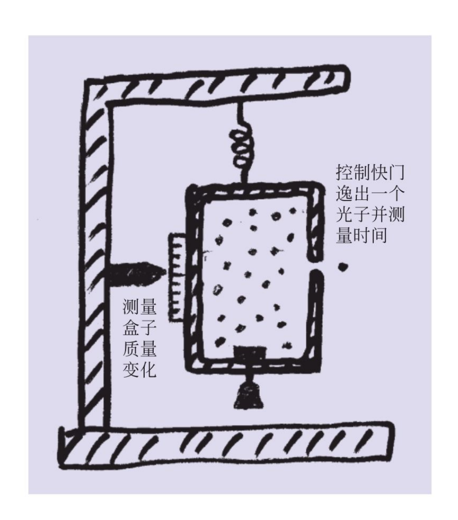



图10-2　爱因斯坦光盒

爱因斯坦抛出这样一个思想实验（图10-2）。假设有一个密封的盒子悬挂在弹簧秤上，盒子里有一定数量的可以辐射光子的物质。一个事先设计好的钟表机构开启盒上的快门，使一个光子逸出，这样，它跑出的时间可被精确测量。同时，由弹簧秤读数可知小盒所减少的质量，这正好是光子的质量，根据相对论质能公式$E=mc^2$，就能算出光子的能量。由于时间测量由钟表完成，光子能量测量由盒子的质量

变化得出，所以二者是相互独立的，测量的精度不应互相制约，这样，时间和能量就能同时精确测量了，因而能量与时间之间的不确定关系不成立。

第7章已经介绍过，除了位置和动量具有不确定关系外，时间和能量也存在不确定关系。如果在某一时刻$t$测量粒子的能量$E$，那么不确定度满足以下关系：

$$\Delta t \cdot \Delta E \geqslant \frac{\hbar}{2}$$

此式表明，在某一时刻辐射出一个光子，如果这个光子的放出时刻确定，它的能量就会有一个很窄的分布范围，不会确定为某一个值；反之，如果光子的能量确定，就不能精确测得光子逸出的时刻。但是现在，在爱因斯坦的光盒实验中，这一规律被打破了，二者均可精确测定，这无异于晴天霹雳，震得玻尔目瞪口呆。



图10-3　第六次索尔维会议上，玻尔和爱因斯坦边走边讨论

玻尔没有马上想出解决之道，他一整天都闷闷不乐。爱因斯坦自信满满，一整天都笑容满面，他相信这一次自己是真的找到了不确定原理的破绽。当时留下的一张照片似乎也暴露了两人的内心（图10-3）。

但是，玻尔也绝非等闲之辈，经过彻夜思考，他终于在爱因斯坦的推论中找到了一处破绽。

第二天，玻尔已经恢复了笑容，他大步流星地走上讲台，在黑板上开始对光盒实验结果进行理论推导，而他用的武器竟是广义相对论的引力时间延缓效应——盒子位置的变化会引起时间的膨胀！经过推导，他竟然导出了能量与时间之间的不确定关系式。玻尔用相对论证明了不确定原理！可以说，不确定原理更让人信服了。

这一回，轮到爱因斯坦目瞪口呆了。尽管台下还有些人对玻尔的反驳提出了质疑，认为他把指针、标尺和光盒当成量子物体来处理并不妥当，但爱因斯坦接受了玻尔的反驳，毕竟，这是思想实验，本来一个光子的质量也不可能从弹簧秤的宏观指针上看出来，所以理论上要读出读数的话，把测量装置看成量子物体并无不妥。

自此以后，爱因斯坦不得不有所退让，承认了哥本哈根学派对量子力学的解释不存在逻辑上的缺陷。但是，他并没有认输，“量子论也许是自洽的，”他说，“但却至少是不完备的。”爱因斯坦改变了策略，他决定不再从外部进攻，但是，他的头脑已经高速运转起来，开始策划从内部攻破这座堡垒。这一等，就是5年，5年之后，爱因斯坦将再次让玻尔陷入无尽的苦恼。

虽然玻尔险胜一招，但是玻尔内心可能并不一定满意自己对于光盒的解释，1962年，玻尔去世，人们发现，他在办公室的黑板上留下的最后一幅图就是爱因斯坦的光盒。这个假想的盒子，也许让他困扰了一辈子。

通过两次索尔维会议的交锋，爱因斯坦败走麦城，哥本哈根学派大获全胜，从此，他们对量子力学的解释被称为量子力学的“正统解释”。

# 第四篇　量子奥义·叠加与测量

# 11 量子第一原理

1930年，狄拉克出版了量子力学的经典教材《量子力学原理》。这是一本不带任何感情色彩的用数学语言写成的物理书，全书没有任何一个图表、一处索引或是参考书目。狄拉克认为，量子世界和人类经历的其他任何事物都不同，如果拿日常行为打比方会产生误导，所以书中几乎没有一处启发性的比喻或类比。

但是，这本书获得了同行们的盛赞。泡利热情地称赞该书是完美之作，尽管他担心这本书写得太抽象以至于太脱离实验观测，但他还是将该书称为“必读的基础教科书”。爱因斯坦也称赞这本书“在逻辑上最完美地呈现了量子理论”。这本书后来终身陪伴着爱因斯坦，他经常在度假的时候将它带在身边作为休闲读物，有时他遇到量子方面的难题时，总是自言自语地念叨：“我那本狄拉克放哪里了？”

在量子力学的理论体系中，从薛定谔方程可以得出一个非常重要的推论——体系的波函数满足叠加原理。因为波函数描述的是体系的量子状态，故称之为态叠加原理。在《量子力学原理》中，狄拉克把“态叠加原理”作为开篇第1章开宗明义，他写道：“量子力学中最基本、最突出的规律之一是态的叠加原理。”在量子力学刚刚建立的年代，狄拉克就认识到了态叠加原理的重要性，量子力学发展到现在，我们不得不佩服狄拉克的眼光。现在，我们可以十分肯定地说，态叠加原理是量子力学区别于经典力学的重要特征，在量子力学中起着统制全局的作用，甚至可以上升到量子第一原理的高度。

对于薛定谔方程，人们发现，它总是齐次线性偏微分方程。这就是说，薛定谔方程的解满足齐次线性微分方程的重要性质——叠加原理。而薛定谔方程的解正是波函数（量子态），于是，就可以得出一条重要的原理——态叠加原理。其表述如下：

如果$\psi_1$、$\psi_2$、$\psi_3$、…、$\psi_n$是某一微观体系的可能状态，那么它们的线性叠加：

$$C_1\psi_1 + C_2\psi_2 + C_3\psi_3 + \ldots + C_n\psi_n$$

也是该体系的可能状态，其中，$C_1$、$C_2$、$C_3$、…、$C_n$是任意常数。

在电子的双缝干涉实验中（见第9章），我们发现，每个电子事实上都在按照波动的特征在运动，自己跟自己发生干涉。这一点我们感觉没法理解，但是了解了态叠加原理以后，就可以发现，态叠加原理正是单个粒子能显示波动性的内在原因。

## 扩展阅读

在高等数学中，齐次线性微分方程有一个重要的性质——叠加原理。叠加原理简单来说，包含两条内容：

（1）齐次线性微分方程的解与任意常数的积也是该方程的一个解；

（2）齐次线性微分方程的多个解的累加也是该方程的一个解。

用数学的语言来表述，就是，对于以下方程：

$$\frac{d^2\psi(x)}{dx^2} + p\frac{d\psi(x)}{dx} + q\psi(x) = 0 \text{（}p\text{ 和 }q\text{ 为实数）}$$

设$\psi_1(x)$和$\psi_2(x)$都是方程的解，那么它们的线性叠加：

$$C_1\psi_1(x) + C_2\psi_2(x)$$

也是方程的解，其中，$C_1$和$C_2$是任意常数（包括实数和复数），称为线性组合系数。

在双缝实验中，从通过某条狭缝的角度来说，电子有两种可能的状态：一种是从狭缝1通过的状态$\psi_1$，另一种是从狭缝2通过的状态$\psi_2$。根据态叠加原理，这两种状态的叠加$\psi_1 + \psi_2$也是电子可能的状态，称之为叠加态。也就是说，按照态叠加原理，电子具有同时穿过两条狭缝的状态（图11-1），因此它可以自己跟自己发生干涉。

如前所述，波函数也叫概率幅（见第5章），由态叠加原理可以发现量子力学和经典物理中计算概率的方法有本质的区别。在经典物理中，如果一个事件可能以几种方式实现，则该事件的概率就是以各种方式单独实现时的概率之和。而在量子力学中，如果一个事件可能以几种方式实现，则该事件的概率幅就是以各种方式单独实现时的概率幅之和（态的叠加），通过概率幅的绝对值的平方才能得知概率。用数学的语言来表述，即

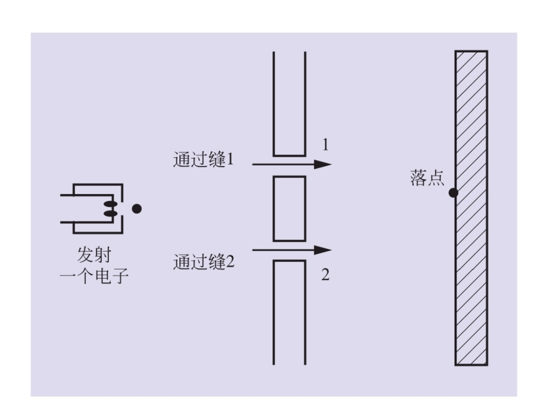



图11-1　电子路径处于通过两条狭缝的叠加态

经典物理：

$$P_{12}=P_1+P_2$$

量子力学：

$$P_1=|\psi_1|^2,\ \ P_2=|\psi_2|^2,$$

$$P_{12}=|\psi_1+\psi_2|^2=P_1+P_2+2\sqrt{P_1P_2}\cos\theta$$

上式中，$\cos\theta$可以描述由于振幅叠加而产生的干涉效应，因此，最后一项可称为相干项。显然，量子概率叠加和经典概率叠加比较，多了一项相干项，这是波动性的直接体现，也是表征叠加态的重要特征参数。

简单来说，经典统计是概率叠加，而量子统计则是概率幅叠加，一字之差，万别千差。量子大师费曼曾经说过，“量子力学里概率的

概念并没有改变，所改变了的，并且根本地改变了的，是计算概率的方法”。

# 12　旋转的硬币

态叠加原理打破了经典物理中描述粒子状态的非此即彼的传统观念，它揭示了量子世界中最重要的性质——叠加态。微观粒子的量子态可以处于各种状态的重叠，这就是叠加态。因为处于多种状态的叠加，所以粒子的某些属性在没进行测量之前是不确定的，只有测量完成后，它的属性才会固定下来，而一次测量只能有一个结果，所以对量子态进行测量，必然破坏其叠加态，导致它变成某一确定态。按照量子力学，这个确定态必然是某一个本征态，虽然可以预测每一个本征态出现的概率，但对于每一次测量，其结果是完全随机的。

例如，电子有一种量子属性叫自旋。自旋并不是电子自身的旋转，它就像质量、电荷一样，是电子的内禀性质。电子自旋可以通过施特恩-盖拉赫实验来测量，如图12-1所示，令电子射线束通过一个不均匀的磁场，电子束在磁场作用下发生偏折，分裂为上下两束，最后在玻璃屏上出现上下两处落点区域。这就说明电子磁矩有两种相反的取向，对应着两种自旋状态，人们称之为自旋向上和自旋向下。

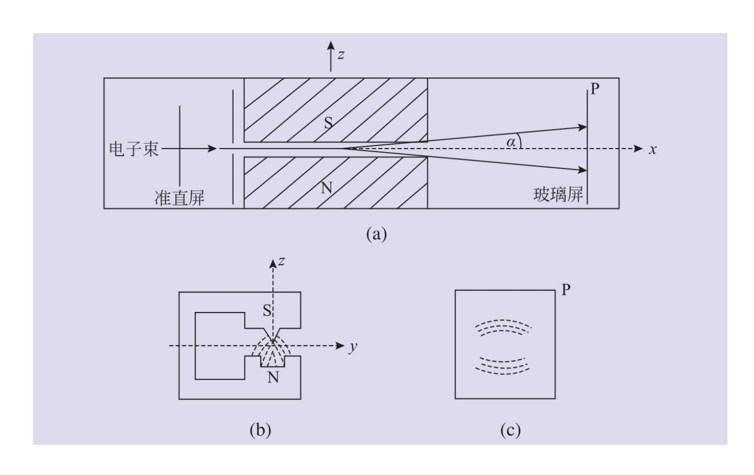



图12-1　施特恩-盖拉赫实验示意图

（a）电子束在不均匀磁场中分裂为两束；（b）不均匀磁场正视图；（c）屏幕落点图像

电子的自旋有两种可能的状态——自旋向上和自旋向下，每个电子都处于自旋向上和向下的叠加态。如果我们用$\alpha$和$\beta$来分别表示自旋向上和向下，那么电子的自旋状态可以记为

$$\Psi = C_1 a + C_2 b$$

这里$\alpha$和$\beta$就是电子自旋的本征态，如果测量一个电子的自旋，那么测量结果可能是$\alpha$，也可能是$\beta$，结果是完全随机的，但是必为$\alpha$和$\beta$其中之一，不会出现其他结果。打个比方来说，叠加态就像一枚旋转的硬币，不管它如何旋转，在你把它拍到桌面上的一瞬间，它不是显示正面就是显示反面（图12-2）。



图12-2　旋转的硬币就像由正面和反面两个本征态组成的叠加态，当你把它拍到桌面上，必然变成正面或反面之一，而不会有其他的结果

如果测量之前你想预测一下测得$\alpha$或$\beta$各有多大的概率，只要看组合系数$C_1$和$C_2$就行了，测得$\alpha$和$\beta$的概率之比是$|C_1|^2$:$|C_2|^2$。如果$\Psi = a + b$，即$C_1 = C_2 = 1$，则测得$\alpha$和$\beta$的概率之比是1：1，即有50%的概率测得自旋向上，50%的概率测得自旋向下。尽管如此，每一次测量的结果都是完全随机的，是不由人为控制的，只有对大量电子进行测量后，才会从统计的角度看到大约有一半电子变成了自旋向上，一半变成了自旋向下。

在经典力学中，我们对物理量的测量是一种旁观者的角度，不会对物理量本身造成任何影响。例如，你想测量子弹的速度和位置，你可以用高速摄像机拍摄子弹的运动轨迹，从而计算它每一时刻的位置和速度。显然，我们不会认为摄像机的拍摄过程会影响子弹的运动，子弹处于一个确定的运动状态，测量过程对测量结果不会造成任何有

实际意义的影响。

但是在量子力学中，叠加态本身就是不确定的，任何轻微的扰动都会对其造成不可逆转的影响，因此，在量子测量中，测量不再是旁观者，而是参与者，测量成了体系本身的一部分。例如，在双缝实验中，如果你不去观察电子穿过哪条狭缝，那么电子处于同时穿过狭缝1和狭缝2的叠加态；而如果你一旦观察到它从哪个狭缝穿过，就相当于完成了一次测量，电子的叠加态就消失了，它变成了从狭缝1或者狭缝2穿过的确定态，于是干涉现象也就消失了。因此，在量子测量中，测量行为和测量结果是关联在一起的。

## 扩展阅读

可以发现，如果$\psi=\frac{\alpha+\beta}{\sqrt{2}}$，测得$\alpha$和$\beta$的概率之比也是1：1，这就说明$\Psi=a+b$和$\psi=\frac{\alpha+\beta}{\sqrt{2}}$代表的是同一种状态。事实上，所有$C_1=C_2$的波函数代表的都是同一种状态，而其中只有$C_1=C_2=\frac{1}{\sqrt{2}}$时正好有$|C_1|^2+|C_2|^2=1$，称之为归一化的波函数，此时测得$\alpha$的概率正好是$|C_1|^2$，测得$\beta$的概率正好是$|C_2|^2$，计算很方便。为了使波函数的物理图像更直观，通常都要求波函数归一化。

纵观量子力学的全部内容，它的成功之处就在于对测量结果的解释和预言，一旦离开了对物理量的观测，它就只剩下一套数学上的演绎与推导。但是，自从量子力学诞生之日起，关于量子测量的解释就是争论的焦点，由此也发展出了多种理论，待后文详述。

# 13 既死又活的猫

“薛定谔的猫”可以说是讨论叠加态的时候人们最喜欢提及的一个例子。作为量子力学的奠基人之一，薛定谔的个人声望使这只猫风靡全世界，但是，很少有人知道，薛定谔当初提出这只猫是为了反驳叠加态。

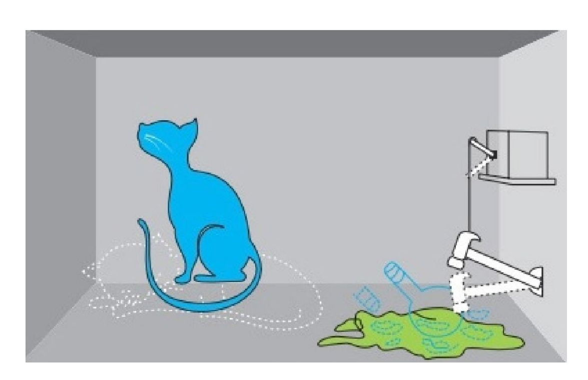


图13-1　薛定谔的猫实验装置示意图

1935年，薛定谔提出一个怪异的思想实验——薛定谔的猫（图13-1）。一只猫被关在箱子里，箱子中有一小块放射性物质，它在1小时内有50%的概率发生一个原子衰变。如果发生衰变，就会通过一套装置触发一个铁锤来击碎一个毒气瓶从而毒死猫。在1小时之内，你无法判断猫是死是活，除非打开箱子看。薛定谔说，按照量子力学规则，可以认为猫处于“死”和“活”的叠加态，只有测量（打开箱子看）才会使它变成确定态。

虽然我们看不到微观粒子的量子态，感觉它十分神秘，但是猫可太常见了，它要么被毒死，要么没被毒死，不管你看不看，它只有这两种可能，怎么可能因为你没去观察就认为它处于“死”和“活”的叠加态？这太荒谬了！就像爱因斯坦说的，月亮在不在天上，与你看不看它有关系吗？

这个实验构建了一种违背常识的叠加态，让人们感觉量子力学好

像十分荒谬。那么，薛定谔为什么要这么做呢？

前文说过，薛定谔虽然自己提出了波函数，但是他却反对波函数的概率诠释，他坚持认为波函数描述的是一种物理实在的波动，不承认叠加态的存在。他和爱因斯坦站在同一条战线上，坚持决定论，反对概率论，所以，他苦心孤诣想出“薛定谔的猫”这个例子，是为了用来反驳叠加态，以此来证明叠加态的荒谬。令人啼笑皆非的是，随着时间的流逝，人们早已忘了他的初衷，还以为他是用这个例子来解释叠加态的。所以，如果你想通过这只怪异的猫来理解叠加态，只能是缘木求鱼，南辕北辙。

抛开薛定谔的初衷不说，就看这个实验，薛定谔的猫到底是死是活？它真的处于“死”和“活”的叠加态吗？事实上，我们前面说过，量子效应只有在普朗克常数的影响不可忽略时才能体现出来，一个宏观物体的宏观性质（“死”或者“活”）根本不可能被普朗克常数影响，根本不具有量子特性，就像一颗子弹的运动根本不具有量子特性一样。在箱子打开前，猫的确存在“死”和“活”两种可能性，但这和“死”和“活”的叠加态是两码事儿。就像你用一粒子弹做双缝试验，子弹存在通过狭缝1和狭缝2的两种可能性，但是你绝对不能说它处于通过狭缝1和狭缝2的叠加态。在双缝实验中，如果你尝试监视电子通过哪条狭缝，将会导致干涉的消失；而在“薛定谔的猫”实验中，如果你尝试监视猫的状态，在箱子里安一个摄像头就行了，并不会对猫的命运造成任何影响，这和叠加态有本质的区别。

反过来，如果你一定要研究猫的叠加态，笔者认为那就要把这只猫所包含的所有粒子的可能性都组合起来，那就是天文数字的叠加和纠缠了，决非简单的“死”和“活”所能描述。

总而言之，从宏观角度来说，不管你看不看，这只猫或者死，或者活，均没有叠加态。所以，薛定谔想要通过它来反驳叠加态是不成立的。



什么是薛定谔的猫？

# 14 环境的力量

在历史上，物理学界对于量子测量的结果并没有争议，但是在如何解释量子测量方面却存在巨大争议。哥本哈根学派提出了“波函数坍缩”解释，他们认为，在一次测量和下一次测量之间，除抽象的波函数以外，这个微观物体并不存在，它只有各种可能的状态；仅当进行了观察或测量，粒子的“可能”状态之一才成为“实际”的状态，并且所有其他可能状态的概率突变为零。这种由于测量行为产生的波函数的突然的、不连续的变化被称为“波函数坍缩”。例如，在电子双缝干涉实验中，每个电子落在屏幕上都是一次波函数坍缩。

但是，根据薛定谔方程演化的量子态，并不会自然地出现波函数坍缩这样的现象，哥本哈根学派找不到具体的细节来说明坍缩的过程。坍缩为何发生？何时发生？持续多久？既然根据薛定谔方程波函数不会坍缩，那么在坍缩的瞬间是什么方程代替了薛定谔方程？于是，这一现象的发生过程成为困扰物理学家们的难题。

还有一个问题，在波函数坍缩之前，粒子有各种可能的状态，在测量的一瞬间，粒子的“可能”状态之一才成为“实际”的状态，并且所有其他可能状态的概率突变为零。爱因斯坦认为，这种瞬间的信息传递是超光速的，是违背相对论的。

总之，波函数坍缩过于突兀，其物理过程的缺失让人怀疑这也许只是人为引入的一种解释实验现象的手段，因此很多人对波函数坍缩的哥本哈根解释并不满意，从而着手寻找新的解释。

一直到20世纪70年代，物理学家们终于找到了一种理论，来说明波函数为什么会坍缩以及如何坍缩的问题，这就是退相干理论。经过几十年的发展，退相干理论已经成为被广泛接受的理论。

由于微观体系具有明显的波粒二象性，所以干涉现象是量子体系最基本的特征。读者还记得双缝干涉实验吗？每一个电子事实上都在自己跟自己干涉，因为它处于通过狭缝1和狭缝2的叠加态。也就是

说，量子叠加态自身具有干涉特性，可以称之为自相干。或者说，叠加态具有相干性。

所谓“退相干”，顾名思义，就是指相干性的退去，表现为波动性逐渐丧失、叠加态逐渐退化为确定态。根据退相干理论，当被测系统与测量仪器和外界环境相互作用后，就会发生退相干过程。如果叠加态的相干效应减弱，称之为退相干；如果相干效应完全消失，称之为完全退相干，此时，量子体系退化为经典体系。退相干是一个客观的物理过程，这一点已经在实验上被多次证明。

观察第11章式（11-2），可以看到：当相干项的数值减小时，相干效应会减弱，发生退相干；如果相干项的数值变为零，则式（11-2）完全退化成式（11-1），相干效应完全消失，退化为经典体系。

还以双缝实验为例，实验本来在真空中进行，但是，如果你想监视电子从那条狭缝穿过，环境中就引入了光子，光子与电子碰撞，就会导致电子的退相干，于是屏幕上的图案就会发生改变。如果光子能量小，会导致电子相干性减弱，此时屏幕上仍然会出现模糊的干涉图像；但如果光子能量大，就会导致完全退相干，干涉图像完全消失。

在量子世界里，干涉现象是普遍存在的，但为什么在经典世界里就观察不到？从量子到经典，是如何过渡的？人们认识到，最初形成的量子观点仅适用于孤立的封闭系统，然而，宇宙中没有任何物体是完全孤立的，因此，不考虑外部环境的作用是不现实的。于是退相干理论提出了这样的观点：自然界中宏观量子效应的缺乏，是由于周围环境造成的退相干效应导致，经典性是量子性退去相干性的结果。

根据退相干理论，相干叠加态只有在与世隔绝的情况下才能够一直维持下去。然而事实上，除了宇宙本身以外，每个真实的系统，不论是量子的或是经典的，都与外部环境密切联系，都是开放的系统。外部环境可以是空气中的分子、原子，也可以是辐射中的光子。它们就像一个个“观测者”，不断地和系统发生耦合作用。这种不可避免的耦合作用会导致系统的相干相位关联不可逆地消失，从而破坏系统的相干叠加性，促使系统的波函数坍缩到某个确定的经典态。

简单来说，一个与环境隔绝的量子系统处于纯态的叠加态，一旦接触外部环境，它与环境的相互作用就将破坏它的叠加态，使系统发生退相干。

退相干理论中有一个参数叫退相干时间，就是体系从量子态演变为经典态的时间，退相干时间与研究对象的大小和环境中的粒子数密切相关。

一个半径为$10^{-8}$ m的分子在空气中的退相干时间约为$10^{-30}$ s；如果把空气抽去，则能延长到$10^{-17}$ s；如果把这个分子放在星际空间，它在那里只能与宇宙微波背景辐射相互作用，估计能延长到30 000年。而对于一个半径为$10^{-5}$ m的尘埃颗粒，由于它太大了（含有大量的内部粒子），即使在星际空间，其退相干时间也只有1 μs。

可见，如果粒子体系足够大，或者环境中有大量粒子存在，退相干时间就会非常非常短。在电子遇到屏幕时，屏幕上的大量粒子会使电子瞬时退相干，于是我们就会测量到一个落点。这就解释了波函数为什么会坍缩。

如果退相干的解释是对的，那反过来想一下，如果把宏观世界的物体与环境完全隔绝开来，是不是就能不退相干，从而处于叠加态呢？基于这样一种推测，物理学家们开始探索跨越经典和量子边界的技术，期望寻找宏观量子叠加态。因为薛定谔猫的名气已经传遍全球，其已经成了宏观量子叠加态的代名词，所以科学家们干脆把具有宏观可区分性的两个或多个态的相干叠加叫做“薛定谔猫态”。采用这个名字只是为了使问题显得更加形象通俗，便于向公众传播，并不是说真的去拿一只猫做实验。事实上，想排除退相干效应是异常困难的，这注定是一条艰难的“寻猫之旅”。

退相干理论为量子世界和经典世界提供了一座桥梁，它可以说明量子行为变迁为经典行为的过程，而且它没有对量子力学的基础表述做任何修改，很多量子实验也证实了退相干现象的存在。因此，退相干理论已经被广泛地接受并应用。现在，在量子计算和量子信息的研究中，如何解决退相干问题已经成为科学家们面对的主要难题。

# 15 人的选择

在1927年的第五次索尔维会议上，狄拉克认为，量子测量结果是自然随机选择的结果，而海森伯则认为它是观察者选择的结果。笔者认为，狄拉克所说的自然随机选择，指的是在可能的结果当中，最终出现哪个结果是完全随机的、不可预测的。而海森伯所说的观察者选择的观点，指的是观察者虽然不能预测观察结果，但是却可以选择观察结果将会有哪些可能性。所以二人的观点并不矛盾。

海森伯的观点，我们可以通过一个简单的例子看出来。

偏振是光的一种特殊性质。自然光可以认为是处于所有振动角度的叠加态，但是，使用偏振片可以将自然光变成某一特定方向的偏振光。当自然光射向偏振片时，可将各个方向的振动分解为平行于偏振方向的振动和垂直于偏振方向的振动，这样，自然光就可看作是竖直偏振光和水平偏振光的叠加。当自然光射过偏振片时，水平偏振光被吸收不能通过，竖直偏振光可以通过，故光强只剩原来的一半，如图15-1所示。

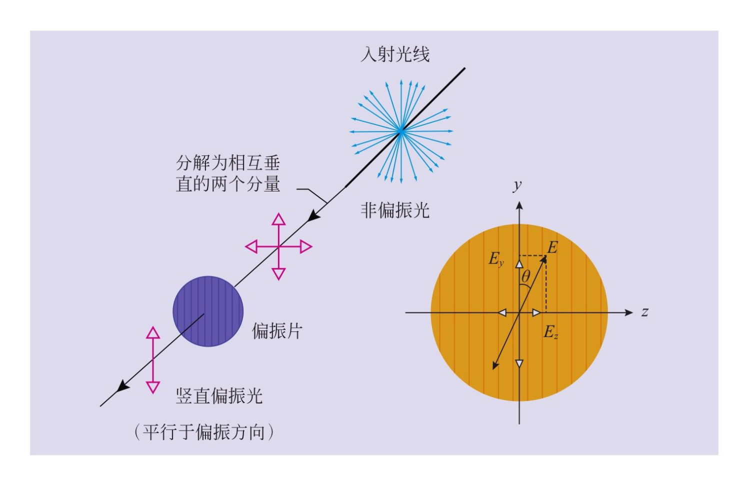



图15-1　自然光包含了所有角度的振动方向，任何一个方向的振动都可以按右下角的方法分解为平行和垂直于偏振方向的振动。当它穿过偏振片时，只

有平行于偏振方向的分量通过，该方向在此处用画在偏振片上的竖线来表示，通过偏振片后就得到了竖直偏振光，光强为入射光的一半

对于单个光子来说也是如此，在它没有通过一个偏振片之前，其偏振方向处于水平和竖直的叠加态，若你进行一次测量，也就是让它射向偏振片，在它接触偏振片的一瞬间，它就会从叠加态变成确定态——或者变成竖直偏振态通过偏振片，或者变成水平偏振态被偏振片阻挡，各有50%的概率。

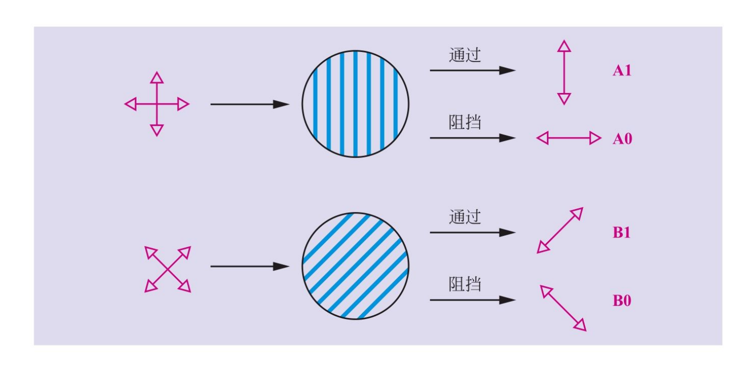



图15-2　偏振片放置角度不同，测量的结果不同（为了表示方便，将图中的四种偏振态记为A0、A1、B0、B1）

于是，偏振片朝哪个方向摆放，就成了对测量结果至关重要的影响因素，而偏振片的摆放角度则完全是观察者选择的结果。如图15-2所示，把两个偏振片放在水平桌面上，一个偏振方向与桌面垂直，另一个与桌面成45°，让光子通过这两个偏振片。显然，光子的叠加态被人为地固定成了十字方向和交叉方向，而测量结果也因此被人为地选择为A0和A1之一或者B0和B1之一。

如果你认为这不算啥，那还有更令人困惑的事情等着你。我们把两个偏振片垂直摆放，光就被完全挡住了，无法通过，见图15-3（a）。但是，如果你在中间加一个45°的偏振片，居然又透光了，见图15-3（b）。

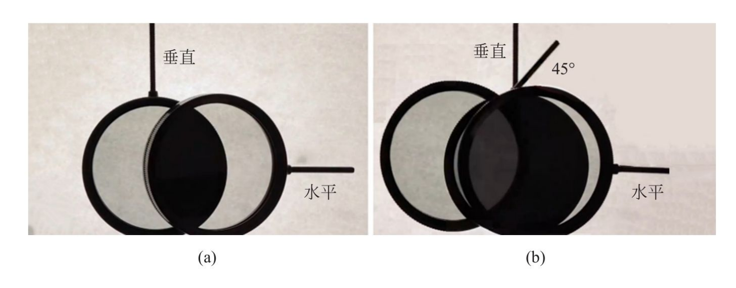



图15-3　偏振片不同堆叠方式的透光效果

（a）两个偏振片垂直摆放；（b）在中间加一个45°的偏振片

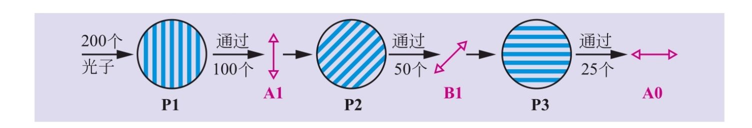



图15-4　让光子连续通过三个偏振片，偏振量子态的变化

下面我们从测量的角度来分析一下上面的现象。如图15-4所示，我们让200个光子依次射向与桌面垂直的偏振片P1，你会看到大约有100个光子通过，于是我们说，这100个光子从叠加态变成了确定态——A1偏振态。这时，如果你在后面再放一个与桌面垂直的偏振片，这100个光子将全部通过，这很好理解，因为它们是确定的A1态。但是，如果你在后面放一个与桌面成45°的偏振片P2，奇怪的事情发生了，大约有50个光子会通过（变成B1态），另外50个被阻挡（变成B0态），这意味着，在通过P2之前，这100个光子可以看作是处于B0和B1的叠加态。

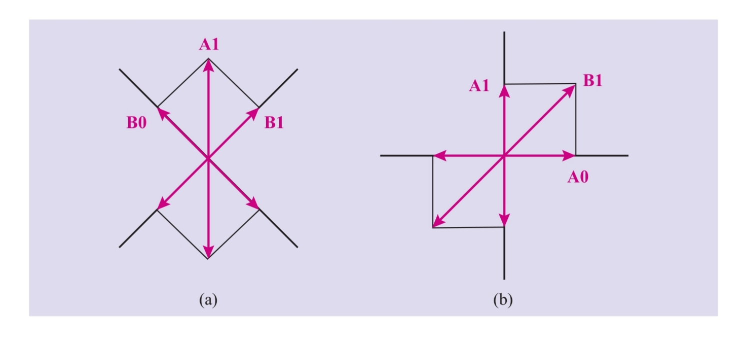



图15-5　光子偏振态在不同基础态下的分解

（a）A1在B0/B1基础态下的分解；（b）B1在A0/A1基础态下的分解

我们可以将A1分解为两个45°方向的振动，从而解释这一现象，如图15-5（a）所示。但是，你仔细想一想，这100个光子到底是确定态（A1）还是叠加态（B0和B1）？如果你不放P2，它就是确定态，如果你放了P2，它就是叠加态。就像海森伯认为的那样，这是观察者选择的结果。

如果在后面再放一片与桌面平行的偏振片P3，这时候，通过P2的50个处于B1态的光子将再次分为两部分，大约有25个通过P3变成A0态，另外25个被阻挡变成A1态，如图15-5（b）所示。也就是说，我们通过后续的两次测量，将原来全部是A1确定态的光子中的一部分变成了A0态！

再继续探究，你会发现，如果你把P2撤掉，是没有光子能通过P3的，但是，加上P2以后，部分光子就能通过P3了，这是不是说明观察者可以选择结果呢？

不可思议是吗？事实就是如此。

费曼在其《物理学讲义》里把以上的结果归纳为量子力学的一条基本原理：任何量子体系可以通过过滤将其分解为某一组所谓的基础态，在任一给定的基础态中，粒子未来的行为只依赖于基础态的性质——而与其以前的任何历史无关。

在上述例子中，A0/A1和B0/B1就是两组基础态。显然，基础态取决于偏振片的方向，而偏振片的方向取决于测量者如何摆放。正是因为测量者可以选择不同的基础态来决定可能的测量结果，人们才开发出了现代量子保密通信的各种方式。关于量子保密通信，留待后文详叙。

## 扩展阅读

### 希尔伯特空间

美籍匈牙利学者约翰·冯·诺依曼（1903—1957）是人人熟知

的“计算机之父”。但是，你可能不知道，他还是一位对量子理论发展做出很大贡献的物理学家。

冯·诺依曼从小就表现出极高的数学天赋，据说他不到10岁就掌握了微积分。1926年，冯·诺依曼获得了布达佩斯大学的数学博士学位，随后来到哥廷根大学，担任数学家戴维·希尔伯特（1862—1943）的助手。这一年，量子力学刚刚建立，是最热门的话题，恰巧海森伯在哥廷根大学举办了一场介绍量子力学的讲座，于是希尔伯特就带着冯·诺依曼一起去听。海森伯讲完以后，令希尔伯特尴尬的是，他竟然没太听明白，只好让冯·诺依曼给他解释一下。冯·诺依曼不但听明白了，还发现了量子力学与希尔伯特最熟悉的数学语言“希尔伯特空间”之间的联系，于是他用“希尔伯特空间”解释了量子理论，希尔伯特这才恍然大悟。

1930年，冯·诺依曼奔赴美国，入职普林斯顿大学。在这里，他把当年的发现写成了一本书，题目叫做《量子力学的数学基础》，于1932年出版。他将“希尔伯特空间”引入量子力学的理论体系，证明了复平面上的向量几何与量子力学系统的各种状态有着相同的公式化特征，从而建立了一整套用于描述神秘莫测的量子现象的数学模型，意义非常重大。就像他在书中说的那样：“由希尔伯特最早提出的数学思想就能够为物理学的量子理论提供一个适当的基础，而无须再为这些物理理论引进新的数学构思。”就这样，冯·诺依曼从一个数学家，摇身一变成为了量子力学大师。现在，“希尔伯特空间”已经成为量子研究者不可或缺的数学工具。

下面我们对希尔伯特空间做一下简单介绍。冯·诺依曼发现，态叠加原理还可以换一种方式来表述：描述体系状态的所有波函数构成一个集合$\{\psi_n\}$，该集合中任意几个波函数的线性叠加可以得到一个新的波函数，这个新的波函数仍然在此集合中。也就是说，该集合对于线性叠加是封闭的，数学上把这样一个集合称为线性空间。如果该集合中所有波函数都已经归一化，则称为希尔伯特空间。

由此，冯·诺依曼得到了一种描述量子体系状态的数学形式——希尔伯特空间。描述体系状态的全部波函数张开一个希尔伯特

空间，量子体系所有的活动都是在这个空间中进行的。

在希尔伯特空间中，一个波函数类似于几何学中的一个矢量，所以波函数也被称为态矢量，简称态矢。

我们以电子的自旋为例，电子的自旋量子态处于自旋向上（$\boldsymbol{\alpha}$）和向下（$\boldsymbol{\beta}$）的叠加态，这样就构成一个二维希尔伯特空间，它是由$\boldsymbol{\alpha}$和$\boldsymbol{\beta}$这两个基矢张成。所有的态矢量都可以用下式来表示：

$$\psi = \cos\frac{\theta}{2}\boldsymbol{\alpha} + \sin\frac{\theta}{2}(\cos\phi + \mathrm{i}\sin\phi)\boldsymbol{\beta}$$

$$(0 \leqslant \theta \leqslant \pi, 0 \leqslant \phi \leqslant 2\pi)$$

这些矢量在球极坐标系中构成一个封闭的球面，球面上每一个点对应的矢量都是一个态矢，也就是说，所有的态矢与球面上的点都是一一对应的，这个球被叫做布洛赫球（图15-6）。

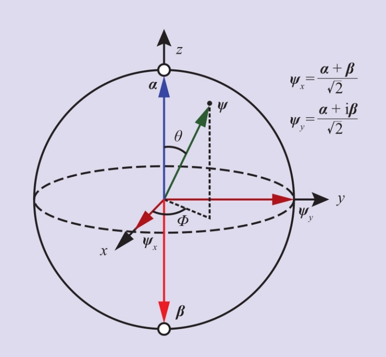



图15-6　布洛赫球

布洛赫球北极和南极的两个矢量分别代表$\boldsymbol{\alpha}$和$\boldsymbol{\beta}$，其他矢量是$\boldsymbol{\alpha}$和$\boldsymbol{\beta}$的不同叠加态（如图中沿$x$轴的矢量$\psi_x$和沿$y$轴的矢量$\psi_y$）。这样，分析波函数（态矢量）的演化就有了一种形象的可视化手段，这在量子计算机的研究中是非常有用的。

现在，波函数已经有了4种叫法——波函数、态函数、概率幅和态矢量。这4种叫法从不同的视角展示了量子力学的特性，其中态矢量主要体现了波函数的叠加特性。不同的形式并不影响有物理意义的结果，所以允许我们选择方便的形式来处理具体的物理问题。

## 扩展阅读

### 中微子振荡与叠加态

中微子是基本粒子之一，宇宙中有大量的中微子，例如，超新星爆发会产生大量中微子，太阳里面的轻核聚变也会放出大量中微子。中微子质量小，不带电，运动速度接近光速，只参与非常微弱的弱相互作用，具有极强的穿透力，能轻松穿透地球，就像宇宙间的“隐形人”。

地球上每平方厘米每秒有600亿~1 200亿个中微子穿过，但是每100亿个中微子中才有一个会与物质发生反应，因此中微子的检测非常困难。1930年，泡利提出了中微子的假说，直到1956年才被观测到，证明了它的存在。

中微子有3味，分别是电子中微子（$\nu_e$）、μ子中微子（$\nu_\mu$）和τ子中微子（$\nu_\tau$），这三味中微子除了质量依次增大外，其他性质完全一样。在对中微子的观测中，人们发现它有一种奇特的性质，就是它能够变身。它在飞行过程中会从一味中微子变成另一味中微子，而且还能变回来，这样不断地变来变去，呈现出周期性的转化，人们称之为中微子振荡。

这三味中微子在理论上是不会相互转化的，那么，为什么在实验中总是观察到周期性的转化呢？

原来，中微子有三种质量本征态（$\nu_1$、$\nu_2$、$\nu_3$），三味中微子都处于这三种质量本征态的叠加态。即：

$$\nu_e = C_1'\nu_1 + C_2'\nu_2 + C_3'\nu_3$$
$$\nu_\mu = C_1''\nu_1 + C_2''\nu_2 + C_3''\nu_3$$
$$\nu_\tau = C_1'''\nu_1 + C_2'''\nu_2 + C_3'''\nu_3$$

三种质量本征态（$\nu_1$、$\nu_2$、$\nu_3$）的波长不一样，这些非常微小的波长差异，在积累了足够长的距离之后，就会变成显著的相位差异，导致在不同距离上叠加态的组合系数不一样。假设一开始我们只有$\nu_e$，经过一段时间，飞行一定距离后，它的组合系数由

$C'_1$、$C'_2$、$C'_3$演化成了$C_1$、$C_2$、$C_3$，变成了

$$\nu = C_1\nu_1 + C_2\nu_2 + C_3\nu_3$$

于是这时候的中微子实际上成了$\nu_e$、$\nu_\mu$和$\nu_\tau$的叠加态，这时候进行测量，就有了$\nu_\mu$和$\nu_\tau$出现的概率，而$\nu_e$出现的概率则比原来低了。随着组合系数的演化，测量结果会周期性变化，这个周期通常在千米量级。

中微子的振荡现象，可以看作是态叠加原理有效性和必要性的直接证据。如果没有叠加态的存在，很难解释中微子振荡现象。

# 第五篇　量子奥义·纠缠

# 16 分而不离

爱因斯坦是量子理论的创始人之一。但他却是坚定的决定论信奉者，他坚信“上帝不会掷骰子”，反对概率论，尤其不认同不确定原理和叠加态。1927年和1930年，他在索尔维会议上两次就不确定原理发起攻击，都被玻尔化解，以失败而告终。但是，爱因斯坦并没有放弃，他改变策略，决定采用反证法，从内部攻破这座堡垒：如果从量子力学基本原理出发，推演出一个十分荒谬的结果，那不就说明量子力学是不完备的吗？

要知道爱因斯坦可是把狄拉克的《量子力学原理》当休闲读物来读的人，虽然大家公认玻尔是量子力学的领军人物，但实际上爱因斯坦对量子力学也是了如指掌。在爱因斯坦苦心孤诣的推敲下，他终于发现了量子力学的一处“破绽”。但是，他这一次不能直接面对面的和玻尔交锋了，因为，欧洲发生了巨大的变化。

1933年，德国纳粹上台，欧洲战云密布。德国国内，排挤犹太人的行动愈演愈烈，连爱因斯坦也未能幸免。犹太人科学家纷纷逃离德国乃至欧洲，在这样的情势下，爱因斯坦于1933年10月移居美国，从此再没回过欧洲大陆。

尽管1933年的第七次索尔维会议仍然按时在布鲁塞尔召开，但爱因斯坦已经没法参加了。而哥本哈根学派以玻尔为首，海森伯、泡利、狄拉克等人悉数出席，可谓阵容强大。虽然持决定论的德布罗意和薛定谔也出席了会议，但失去了主心骨爱因斯坦，两人都没有向量子力学提出挑战，这令玻尔大大松了一口气。看起来，论战似乎已经尘埃落定。

尽管这时候玻尔和海森伯还是为了量子力学而并肩战斗的情同父子的师徒，但他们不会想到，几年后，两人的关系会发生急剧的变化。1940年，德国出兵占领了丹麦，玻尔的研究所被监控起来。1941年，海森伯造访玻尔的研究所，这时候，他已经成为德国原子弹计划的总负责人，身份特殊。两人的谈话是在玻尔的办公室里进行

的，海森伯试探性地问了些问题，而玻尔假装没听懂，于是这次简短的谈话不欢而散。1943年9月，玻尔秘密逃离丹麦，直到战后才回来。战后，海森伯被美军俘虏，送到英国关押起来，但在1946年被释放。后来，海森伯和玻尔又有几次见面，但两人的关系再也不可能回到从前了。

1933年10月，爱因斯坦移居美国后，入驻普林斯顿高等研究院，这一年，他已经54岁了。他原来习惯说德语，但在这里需要靠英文交流，由于他的英文说的不太好，所以第二年，他给自己招了两名助手。一个是来自麻省理工学院的25岁年轻人罗森(Rosen)，另一个是在俄罗斯出生的39岁的波多尔斯基(Podolsky)。

生活步入正轨后，爱因斯坦终于开始实现他再次挑战量子力学的计划了。他把自己发现的量子力学的“破绽”告诉他的两个助手，讨论了几次以后，爱因斯坦安排罗森做大部分的数学计算，然后安排波多尔斯基执笔撰写论文。

1935年年初，波多尔斯基将论文写好了，爱因斯坦看过之后不太满意，但他也不想再改了，因为毕竟需要用英文撰写，这对习惯用德语的爱因斯坦来说并不容易，所以他最后同意发表。1935年5月，这篇题为《量子力学对物理实在性的描述是完备的吗？》的论文发表在《物理评论》上，这篇论文的观点后来以三位作者姓名的首字母命名，被人们称为“EPR佯谬”。

爱因斯坦随后写信向他的坚定追随者薛定谔介绍了论文的由来：“因为语言问题，这篇论文在长时间的讨论之后是由波多尔斯基执笔的。我的意思并没有被很好地表达出来。其实，最关键的问题反而在研究讨论的过程中被掩盖了。”即便如此，EPR论文中的观点仍然引起了量子力学界的震动。“太不可思议了！”“这怎么可能？”是很多人的第一反应。薛定谔也备受鼓舞，忍了这么多年，终于要扬眉吐气了，他不但给EPR论文中描述的粒子状态起了个名字叫“纠缠态”，还趁势提出了“薛定谔的猫”佯谬（见第13章），为担任主攻的爱因斯坦输送炮弹。

薛定谔还是有两下子的，“纠缠态”这个名字起的非常形象，

比“EPR佯谬”好记多了，一下子就传开了。那么，到底什么是纠缠态呢？

爱因斯坦发现，根据量子力学原理可推导出一个结论——对于一对出发前有一定关系、但出发后完全失去联系的粒子，对其中一个粒子的测量可以瞬间改变任意远距离之外另一个粒子的状态，即使二者间不存在任何连接，这对粒子就处于“纠缠态”。这个改变在瞬时发生，不需要任何传递时间，也就是说，这个改变是超光速的。

设想有一个量子系统由两个电子A和B构成，但两个电子的总自旋为零，这意味着它们总是处于自旋相反的状态。现在将两个电子分别置于相距遥远的两个地方，例如，A在地球上，B在火星上。按照量子力学，这时候每个电子都处于自旋向上和自旋向下的叠加态，是不确定的。但如果对地球上的电子A进行测量，假设其随机变为自旋向上的确定态，那么火星上的电子B会瞬间变成自旋向下的确定态，即使你没对它测量（图16-1）。也就是说，B的状态似乎是瞬间被A的测量所控制，这种控制行为以超光速的方式发生。这是从量子力学原理推演出来的必然结果。

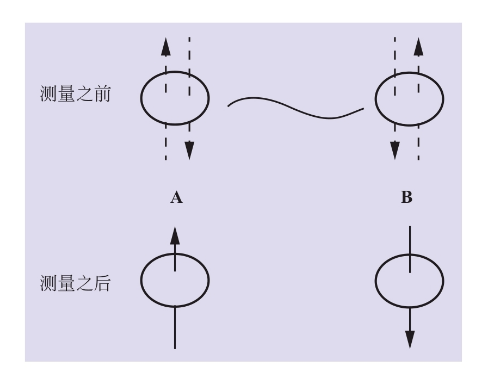



图16-1　纠缠态电子测量前后自旋状态的变化

根据量子力学，处于纠缠态的粒子，即使空间上分离遥远，仍然存在内在量子关联，它们的量子关联与距离无关，对其中一个粒子的任何操作都会瞬时地改变另一个粒子的状态。爱因斯坦抓住的“破绽”，就是这个“瞬时改变”，他认为这违反了相对论里信息传递速度不能超过光速的原理，将违背因果律，所以是根本不可能的。为了凸显

其“荒谬”，爱因斯坦把它叫做“幽灵般的超距作用”，以此来证明量子力学是不完备的。

EPR论文一经发表，哥本哈根学派就坐不住了。泡利给海森伯写信说：“爱因斯坦再一次公开抨击量子力学，甚至发表在5月15日的《物理评论》上，还有波多尔斯基和罗森也跟着起哄。正如我们都知道的，这种事情无论何时发生，都是一场灾难。”然后泡利鼓动海森伯立刻撰文反驳。

## 扩展阅读

上述例子如果用量子力学的语言来描述，可以这样表述：电子的自旋量子态处于自旋向上（设其波函数为$\alpha$）和向下（设其波函数为$\beta$）的叠加态，两个电子A和B总自旋为零的状态只有两种可能：A上B下（$\alpha_A\beta_B$）和A下B上（$\beta_A\alpha_B$），因此，AB系统的状态应当是

$$\psi_{AB}=\frac{1}{\sqrt{2}}(\alpha_A\beta_B+\alpha_B\beta_A)$$

或

$$\psi_{AB}=\frac{1}{\sqrt{2}}(\alpha_A\beta_B-\alpha_B\beta_A)$$

$\psi_{AB}$就是纠缠态（式中，$\frac{1}{\sqrt{2}}$为归一化系数）。如果进行测量，系统将有50%的概率坍缩为$\alpha_A\beta_B$，在此态中A电子自旋向上B电子自旋向下；另有50%的概率坍缩为$\beta_A\alpha_B$，此状态中情形刚好相反。由此不难看出，无论测量使系统坍缩到哪个状态，两电子的自旋方向总是相反。虽然你无法预测单次测量结果，但是你能确定，无论A变成什么，B总是与它相反。

海森伯还没想好怎么反驳，玻尔就已经行动了。当玻尔看到EPR论文后，大惊失色，他立即放下手头的一切工作来思索如何应对。经过3个月的艰苦工作，玻尔终于把回应提交给《物理评论》杂志。他

的论文题目和EPR论文题目一模一样：《量子力学对物理实在性的描述是完备的吗？》。

实际上，玻尔的反驳并不像前两次那样有力，因为“纠缠态”的推论本来就没有错，玻尔也承认这种推论结果的存在，不过，爱因斯坦认为这种结果根本不可能发生，而玻尔认为是可以发生的，仅此而已。也就是说，对于论文题目，爱因斯坦给出的答案是“否”，而玻尔给出的答案是“是”。

这样的争论其实陷入了哲学上的争论，是不会分出胜负的。狄拉克最开始被爱因斯坦震住了，他对身边的人说：“现在我们不得不重新开始了，因为爱因斯坦证明量子力学行不通。”不过，当玻尔的回应发表后，他又改变了主意，他选择相信玻尔，因为量子力学早已证明它的价值，没必要推倒重来。但是，他的哲学观点还是动摇了，他后来在1975年的一次演讲中说道：“关于现在的量子力学，存在一些很大的困难……我认为很可能在将来的某个时间，我们会得到一个改进了的量子力学，使其回到决定论，从而证明爱因斯坦的观点是正确的。”

从狄拉克态度的变化，就能看出爱因斯坦这一次对量子力学的反击是相当有力的，没有人认为超光速的变化是可能的，除了玻尔。很显然，要想一分胜负，只有通过实验来判定。可惜的是，纠缠态实验太难做了，玻尔和爱因斯坦都没有在有生之年看到它，这真是物理学界的一大憾事。而这也导致爱因斯坦一辈子都不接受量子力学对世界本质的描述。海森伯在回忆文章中写道：“1954年，爱因斯坦去世前几个月，他同我讨论了这个问题。那是我同爱因斯坦度过的一个愉快的下午，但一谈到量子力学的诠释时，仍然是他不能说服我，我也不能说服他。他总是说：‘是的，我承认，凡是能用量子力学计算出结果的实验，都是如你所说的那样出现的，然而这样的方案不可能是自然界的最终描述。’”

1955年4月18日，爱因斯坦逝世，享年76岁。1962年11月18日，玻尔逝世，享年77岁。虽然玻尔的黑板上仍然画着爱因斯坦的光盒，但失去了主角的世纪论战已然成为了历史的绝唱。



什么是纠缠态？爱因斯坦和玻尔争论的焦点是什么？

# 17　扔掉骰子

关于量子纠缠，有一个常见的错误比喻，这就是爱因斯坦提出来的手套比喻。爱因斯坦认为，一对纠缠的粒子在出发前其实已经固定了状态，不过是你不知道罢了。就像把一双手套分别放在两个密闭的箱子里，当你打开一个箱子发现是左手的时候，你瞬间就知道另一个是右手。这样的话就不存在超光速的问题了，因为根本没有信息传送。但事实上这个比喻是不符合量子力学的思想的，因为这里否定了叠加态的存在，这是爱因斯坦所支持的决定论思想的体现。如果按量子力学的思想，两个箱子里的手套都是处于左右手套的叠加，是不确定的。如果还用两个总自旋为零的纠缠态电子为例，那么就是片面的。爱因斯坦对于纠缠与测量的看法可以用图17-1表示，它与图16-1是完全不同的。

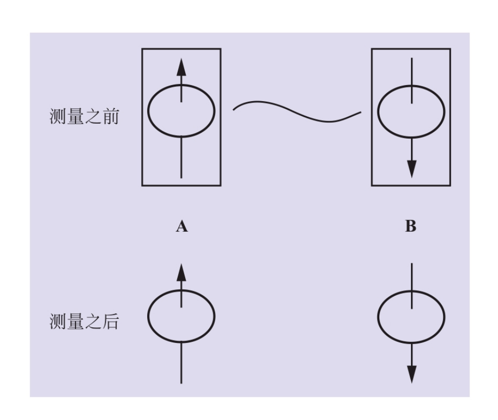



图17-1　决定论者认为的纠缠与测量

事实上，爱因斯坦的观点在那个年代是不好反驳的。因为量子力学虽然承认叠加态的存在，但是你是没法直接观察叠加态的，因为你一测量它就变成了确定态。所以即使之前是左右手套的叠加，但是你一打开箱子只能看到是左手或右手，和爱因斯坦看到的结果是一样的。所以玻尔和爱因斯坦都无法证明对方是错的。

持决定论的物理学家们认为，目前量子理论之所以是一个概率统计理论，是因为还存在着尚未发现的隐藏变量，简称为“隐变量”。如果能找出这些隐变量并把它们加入量子力学的方程里，就可以对微观粒子的运动状态作出“精确”的描述，而不只是“概率”性的描述。这种理论被统称为隐变量理论。在爱因斯坦的支持下，这样的思想一直没有绝迹，虽然势力弱小，但一直坚持与量子力学的正统解释做对抗。

最早的隐变量理论就是德布罗意的“导波理论”。在导波理论中，德布罗意认为，粒子和波是同时存在着的，粒子就像冲浪运动员一样，乘波而来，在波的导航下，粒子从一个位置到另一个位置，它是有路径的。但是，在第五次索尔维会议上，他被泡利批驳得哑口无言，爱因斯坦也没有给予他相应的支持，这让德布罗意非常失望。几天后会议结束，爱因斯坦要回家了，也许是出于歉意，他拍着德布罗意的肩膀说：“要坚持，你的路子是对的。”但是爱因斯坦的鼓励并没有起到作用，德布罗意放弃了他的理论，没有继续往下研究。

随着量子力学的蓬勃发展，隐变量理论陷入低谷，爱因斯坦虽然在不断给量子力学挑毛病，但他自己并没有提出一个隐变量理论。

直到20世纪50年代，隐变量理论才重新焕发出生机。1951年左右，在爱因斯坦的鼓励下，普林斯顿大学的物理学家戴维·玻姆（1917—1992）将德布罗意的导波理论重新挖掘出来并加以修葺，发展出一个新版本的隐变量理论，被称为玻姆理论，也叫量子势理论。他在1952—1954年期间接连发表数篇重要论文，奠定了量子势理论的基础。



玻姆

在玻姆理论中，波和粒子同时存在，粒子沿着一条由波函数控制的确定轨迹演化。这套理论能让人们用类似牛顿力学的方法来研究量子世界的规律，而且还能解释许多量子实验，这让持决定论思想的人们大为振奋。

玻姆的父亲原籍奥匈帝国，后迁居美国，玻姆在美国出生。1947年，玻姆获得博士学位，来到普林斯顿大学成为助理教授，担任量子力学课程的教学。1951年，他撰写了《量子理论》一书，由于这部书清晰地阐述出量子力学公式背后的重要物理思想，并很详细地讨论了通常容易被忽视的困难问题，如量子理论的经典极限、测量问题以及EPR佯谬等，一出版便大受欢迎。在书中，玻姆清醒地指出，EPR佯谬中所揭示的量子纠缠关系，是一种“非因果关联”，即使存在这种超距作用，也不会破坏因果律。所以EPR佯谬对量子理论的杀伤力，并没有爱因斯坦想象的那么大。

在完成《量子理论》一书后，玻姆将他的书分寄给了爱因斯坦、玻尔和泡利。玻尔没有答复。泡利热情地称他写得好。爱因斯坦最为重视，近水楼台先得月，他直接邀请玻姆到他寓所进行讨论。玻姆找到爱因斯坦，与他做了详尽的讨论。在与爱因斯坦的讨论中，玻姆极大地强化了这样一种信念，就物理学应该对物理实在作出客观而完备

的描述而言，量子理论缺少了某种基本的东西。于是，在爱因斯坦的鼓励下，他开始发展隐变量理论。

不久，玻姆在普林斯顿大学的合同期满，当时麦卡锡主义盛行，他害怕在美国遇到不测，于是在1951年秋离开美国到巴西任教。果然不出所料，玻姆在巴西期间，美国官方取消了他的护照，致使玻姆开始了流亡国外的学术生涯。1955年，玻姆辗转到了以色列，1957年又移居英国。多年的漂泊并没有扼杀他的研究热情，他在以色列期间对EPR佯谬进行了详细的逻辑梳理和重新阐释，为后来贝尔发现著名的贝尔不等式做出了重要铺垫。

在玻姆的理论中，波函数被重新解释为一种表达客观实在的场。玻姆假设存在一种实在的粒子，其运动嵌在场中，沿着实在的空间轨道，并且依照强加的“制导条件”，“受制”于相位函数。于是，每一个场中的每一个粒子均具有精确定义的位置和动量，沿着相应相位函数决定的轨道运动。这样得到的运动方程不仅依赖于经典势能，还依赖于由波函数决定的另一种势能，玻姆称之为量子势。

按量子势理论，原则上我们能追踪每一个粒子的轨迹。但是由于我们无法确定每个粒子的初始条件，所以才只能计算概率。概率仍然联系着波函数的振幅，但这并不意味着波函数只有统计意义。相反，波函数被假设具有很强的物理意义——它决定了量子势的形状。图17-2给出了对具有一定量子势初始条件的电子计算出的双缝实验的运动轨迹。可以看到，各条轨迹在离开每一缝隙后立刻发散，但它们互不相交。两个缝隙的轨迹在正中间有分界线，各占屏幕的一半。电子会沿图中的某一条轨迹运动，然后落在屏幕上，每个电子有不同的初始条件，所以它们各自沿着不同的轨迹到达屏幕，总的结果是屏上的干涉图像的形成。

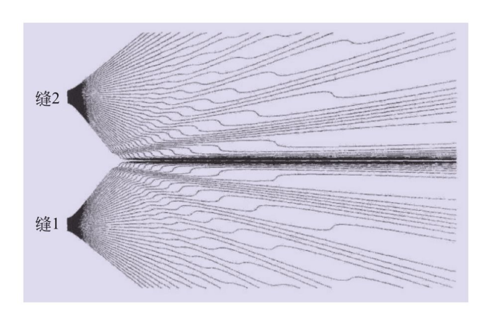



图17-2　用量子势计算出的电子通过双缝的理论轨迹

量子势理论虽然认为粒子的位置和动量在原理上是可以精确确定的，但也承认测量仪器或测量过程对波函数有重要影响，因而会直接影响量子势，从而影响粒子路径。所以测量仪器仍然是关键，量子粒子的轨迹取决于实验设置。在测量仪器对测量结果有决定性影响这一点上，玻姆理论与玻尔的主张实际并不冲突。

“整体性”是量子势理论的核心，量子势实际上将空间里的所有东西看作一个不可分割的整体，任何测量仪器的变化都将导致整个量子势场的变化。量子势理论采取的是“自上而下”的方法：整体比其局部之和具有大得多的意义，并且实际上决定着各个局部的性质和行为。

到了20世纪80年代，玻姆又将其理论进一步发展，提出了“隐缠序理论”。他认为，物理世界有确定的秩序，不过这些信息因为波函数“卷起”而隐藏，一切可被感知和加以实验的特征（显析序）乃是包含在隐缠序里的潜在性的实现，此时波函数被“展开”。隐缠序不但包含这些潜在性，而且决定着哪一个将被实现。在此，波函数的卷起和展开活动是最基本的。波的性质和粒子的性质在波函数不断地卷起和展开中得到体现。

1992年，玻姆逝世，从1952年提出理论到1992年逝世，在40年的时间里，除了寥寥几位物理学家的支持，玻姆几乎一直都是在孤独地耕耘着这片土地。但是，在玻姆去世以后，玻姆理论受到越来越多的关注，陆续出现了一些研究将玻姆理论继续向前推进。有的学者将其推广到相对论时空中，有的学者打通了玻姆理论与量子场论的联系。尽管前路艰难，但这是决定论者眼里一点微弱的希望之光，它能

否成功还需要后来者继续探索。

# 18　不等式的判决

爱因斯坦和玻尔的世纪之争，在20世纪60年代终于迎来了转机，人们终于能够将争论从哲学层面转移到物理上来，这其中最大的功劳，要归功于贝尔不等式的发现。

约翰·斯图尔特·贝尔（1928—1990）在1956年取得了英国伯明翰大学的物理学博士学位。他先后在英国原子能管理局和日内瓦的欧洲核子研究组织（CERN）工作。虽然他的主业是从事粒子物理学和粒子加速器的研究，但他的“业余爱好”是探索量子论的基本问题。上大学时，贝尔的物理成绩非常优秀，但他不满意老师讲授的量子论，他发现量子论某些神秘的特性在课堂上没有得到解释，因此，他一直想找到答案，所以就利用业余时间自己进行研究。

1963年，贝尔休了一年假，离开欧洲去美国访学。他终于有时间全身心地投入自己的“业余爱好”当中，去真正探索量子力学的核心问题。1964年，他回到欧洲，连续写了两篇论文。正是这两篇论文，让他的“业余爱好”成了他主要的物理学成就。

第一篇是《论量子力学的隐变量问题》，这篇论文勇敢挑战了冯·诺依曼，指出了他关于隐变量理论的错误论断。冯·诺依曼在《量子力学的数学基础》一书里，假设几个可观测量之和的预期值等于其中每一个可观测量的预期值之和，并由此证明能够减少量子体系不确定性的“隐变量”是不存在的。贝尔指出这一假设从物理学角度看是不成立的，这样，冯·诺依曼否定隐变量理论的论断就是错误的。这一发现，消除了物理学界对隐变量理论多年的误解。

第二篇论文题目叫《论EPR佯谬》，在这篇论文中，贝尔提出了著名的贝尔不等式。

贝尔发现了当年玻爱论争中的一个重要事实：所谓“EPR佯谬”根本不是什么佯谬。爱因斯坦和玻尔争论的焦点就在于纠缠态可不可能存在。他发现，纠缠态跟爱因斯坦所坚信的定域关联无法并存，但

是，如果是非定域关联，纠缠态就是可以存在的。

在两个空间上分离的物理系统中，对一个系统的作用（如测量）不会立即对另一个系统产生影响，这就叫“定域关联”。定域关联建立在一系列从一点到下一点、在空间连续传递的影响机制之上，所以一定时间内，因果关系只会维持在特定的区域，影响速度不能超光速。但是“非定域关联”就不受光速的限制，对一个系统的作用会瞬间对另一个系统产生影响。

所以说，问题的关键是如何找到一个可行的实验方案，使定域关联和非定域关联的实验结果具有明显的区别，这样就能判断谁是谁非了。经过仔细研究，贝尔终于推导出一个计算关联程度的不等式，如果是定域关联，就满足这个不等式，如果是非定域关联，就违背这个不等式。这就是贝尔不等式。

按照贝尔不等式，如果两个纠缠态粒子出发后就确定了状态（爱因斯坦的观点），那么，这两个粒子的测量结果关联度：

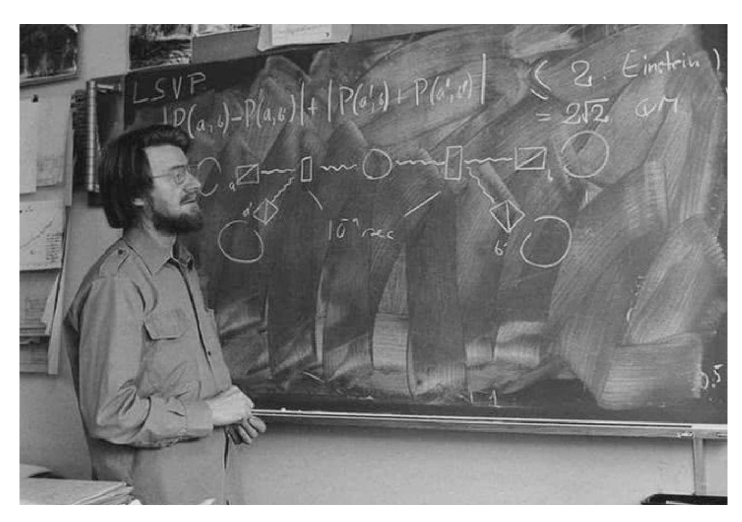
$$|S| \leq 2$$

反之，如果两个纠缠态粒子出发后状态不确定，只有测量时才会随机变化（量子力学的观点），这个关联度将会突破2，最大达到$2\sqrt{2}$。

贝尔不等式给出了定域和非定域的检验标准，具有重要意义，但是，不知为何，贝尔的文章没有发表在知名期刊上，而是刊登在名不见经传的美国《物理》杂志上。这个杂志1964年刚刚创刊，贝尔的文章就发表在创刊号上。更悲催的是，《物理》杂志竟然没办下去，只发行了一年就停刊了，成为历史上最短命的物理学杂志。于是，贝尔不等式一度被埋没在浩瀚的文献资料里，不为人知。

1967年，美国哥伦比亚大学的博士研究生约翰·克劳瑟在图书馆查阅资料时，偶然翻阅到了贝尔的论文。克劳瑟读完这篇论文，马上意识到，贝尔不等式可以验证EPR佯谬的实质。克劳瑟的博士研究课题是宇宙微波背景辐射和射电天文学，但他和贝尔一样，“业余爱好”也是量子理论。克劳瑟对EPR佯谬非常熟悉，也很了解波姆的隐变量理论，自己平时没事就思考这个问题，所以当他看到贝尔不等式以

后，马上茅塞顿开。他决定自己做实验，验证贝尔不等式是否成立。




贝尔在讲解贝尔不等式

作为一个学生，克劳瑟自己并没有独立的研究经费，导师也对于他的研究计划不感兴趣，所以他只好自己东拼西凑收集实验器材，设计实验方案，准备的非常艰难。1969年，他终于完成了实验方案设计，于是给一个物理研讨会寄去了论文摘要，介绍验证贝尔不等式的实验可以如何设计。这一摘要发表在1969年春美国物理学会华盛顿会议的《快报》上。

不久后，孤军奋战的克劳瑟接到了一个电话，是波士顿大学的阿伯纳·西摩尼和迈克尔·霍恩打来的。克劳瑟与他们并不相识，但是，当他们表明来意后，克劳瑟激动起来，他们要跟克劳瑟合作研究！原来，西摩尼和霍恩也在筹划相同的实验内容，当他们看到克劳瑟的会议摘要后，就联系了他。两人已经找好了实验场地，还找到了一位实验物理学家理查德·霍尔特帮忙，这对克劳瑟来说简直是雪中送炭，他一分钟也没多想，立刻答应下来。就这样，他们一起投入了这项研究。

很快，他们就完成了一篇开创性的论文，将其发表于1969年的《物理评论快报》。该文取消了贝尔不等式的一条特殊的限制性假设，从而改良了贝尔不等式，使它的判定结果更加可靠，实现了新的

理论突破，并详细描述了如何用一个改进的实验来验证贝尔不等式。

随后，他们开始开展实验。首先，他们需要一对处于纠缠态的粒子，他们选择了“孪生光子”。对于某些特殊的激发态原子，电子从激发态经过连续两次量子跃迁返回到基态，可以同时释放出两个沿相反方向飞出的光子，而且这个光子对的净角动量为零。这种光子称为“孪生光子”。他们选择了钙原子（$^{40}Ca$），将其用强紫外线激发后，会放出“孪生光子”，如图18-1所示。

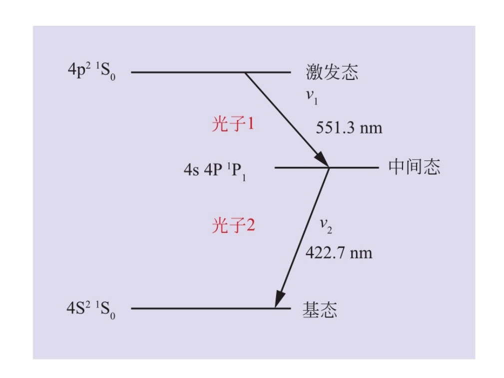



图18-1　钙原子放出“孪生光子”能级变化图

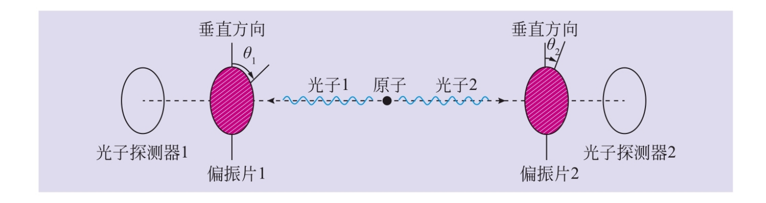



图18-2　通过处于纠缠态的孪生光子检验贝尔不等式的实验示意图

孪生光子产生后沿相反方向飞出，已经没有任何联系，但是因为它们的净角动量为零，所以从理论上来讲，如果你对其中一个光子进行偏振方向测量，另一个光子就必须得和这个光子保持偏振方向一致，否则就没法维持净角动量为零。这就说明，这两个光子是相互纠缠的。

实验示意图见图18-2。把两块偏振片分别放在左右两边，分别测量这对“孪生光子”的偏振方向，然后计算偏振关联度。为了验证贝尔

不等式是否成立，需要改变两个偏振片的夹角，让它们的夹角在-90°～90°的范围内任意变化。量子力学和隐变量理论之间的差别非常微小，研究者只有精确地测量光子对在不同偏振角度下的偏振关联度（图18-3），才能判断哪一种理论是正确的。

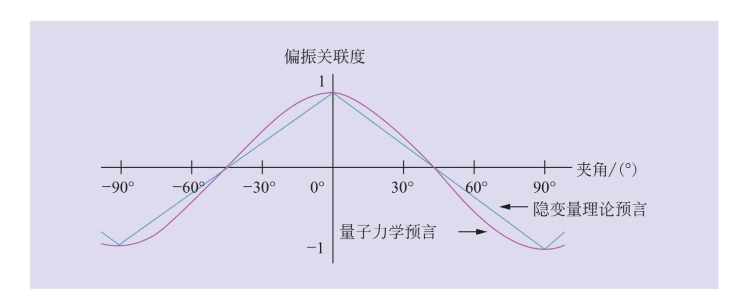



图18-3　量子力学和隐变量理论预言的偏振关联度曲线

由于他们使用的光信号很弱，还有很多杂散的非相关的光子，所以实验难度很大。但是经过艰苦的努力，他们终于获得了有效的实验数据，结果是：贝尔不等式不成立！这一结果有力地支持了量子论，否定了爱因斯坦的定域实在观，从而证实非定域关联确实存在。

克劳瑟等于1972年发表了实验结果，但是也遗留了一些问题。受条件限制，他们的实验中有大量未被观察的光子，实验所用的探测器功效也十分有限。因此，探测器的有限功效和大量未被观测的光子对实验结论的影响究竟有多大，就成为一个重大问题。

随着技术的进步，激光技术被引入到实验中，人们有了完善这一实验的能力。1982年，法国巴黎大学的阿莱恩·阿斯派克特采用激光激发钙原子，又做了一系列精度更高、实验条件更苛刻的实验。他把两个偏振片之间的距离增加到13 m，采用了声光调制器控制的量子开关变换偏振片方向。光传输这13 m的时间是43 ns，而量子开关时间仅有6.7～13.3 ns，这样就排除了两个光子在进入开关前有相互联系的可能。为了消除实验的系统误差，他们还采用了光子双通道的方案，使光子先经过一道闸门，然后进入偏振器，闸门可以改变光子的方向，引导它去向两个不同的方向。最后把四个通道的测量数据汇总到监测器中进行符合处理（图18-4）。

阿斯派克特的实验思想之精巧，设计之精密，设备器材之精良堪称一绝，使业内人士无不赞叹。最终，他以极高的精度确切地证明了贝尔不等式不成立，更关键的是，实验数据与量子理论符合的很好（图18-5），隐变量理论输给了量子力学。

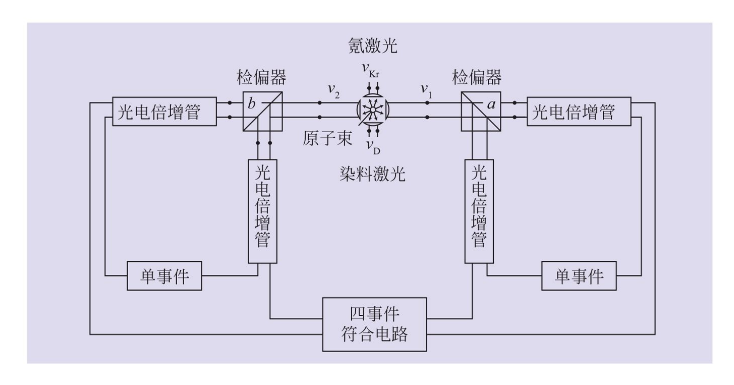



图18-4　阿斯派克特检验贝尔不等式的实验装置图

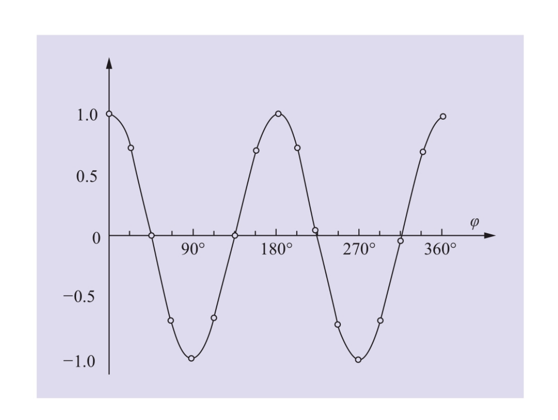



图18-5　阿斯派克特的实验结果

这一实验是阿斯派克特的博士论文研究成果。1983年，在阿斯派克特的博士论文答辩会上，贝尔亲自到场，考察了他的实验，对这一成果赞叹不止。从1964年贝尔不等式发表，到1982年阿斯派克特实验成功，物理学家们经过近20年的奋斗，终于为玻爱之争找到了答案。对贝尔来说，也终于了却了他的一桩心愿。

到了20世纪90年代，人们把这个实验中两个偏振片之间的距离增

加到近11 km，结果仍然没变。而且可计算出光子做出反应的速度至少超过了光速的1000万倍！这个结果证实了爱因斯坦所不喜欢的“幽灵般的超距作用”确实存在，给予了非定域关联以绝对的支持，定域关联被彻底否定。

尽管现在绝大多数人已经承认了量子力学的胜利，但是隐变量理论并没有完全认输，爱因斯坦的支持者仍然从极为苛刻的角度指出上述实验仍然存在“漏洞”。所以直到现在，物理学家们还在进行着条件越来越苛刻的实验。但是，无一例外，实验越精确，结果与量子力学符合的越好。而隐变量理论也在积极求变，玻姆的理论就发展成了非定域的隐变量理论，并不能被贝尔实验排除（非定域的隐变量已经悖离了爱因斯坦的初衷，爱因斯坦坚持的是定域关联）。总之，这场争论还没有完全落幕。

双粒子纠缠现象从实验上被证实以后，人们自然而然地想到了多粒子纠缠的可能性。

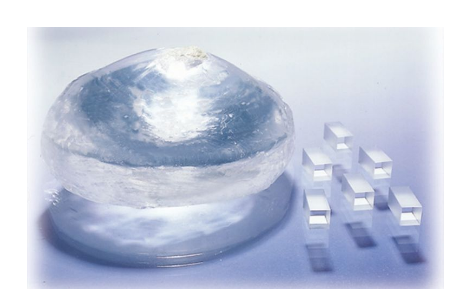


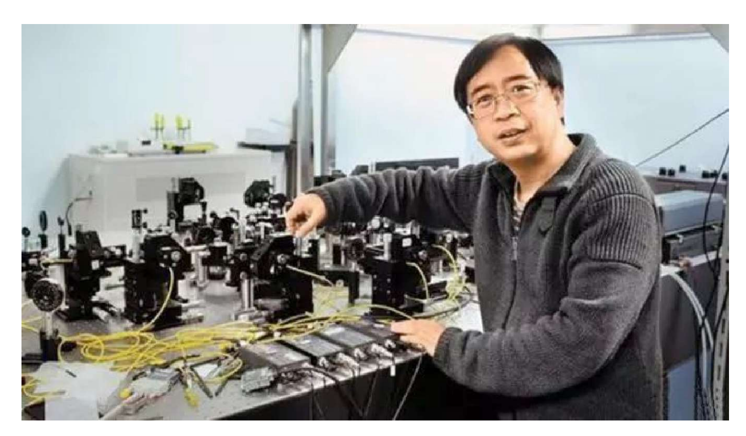
图18-6　BBO晶体

（https://gb.castech.com/product/BBO-%E6%99%B6%E4%BD%93-106.html）

1983年，中国科学院福建物质结构研究所发明了一种性能优异的非线性光学晶体——BBO晶体（偏硼酸钡晶体，图18-6），其很快在量子光学领域获得了广泛应用。1995年，奥地利物理学家塞林格（Zeilinger）团队发明了利用BBO晶体来实现双光子偏振纠缠的方法——用一个紫外激光脉冲照射BBO晶体，可以有一定概率产生一对偏振方向相互垂直的纠缠光子对。这为多光子纠缠的制备提供了基础。

那么如何制备三个相互纠缠的光子呢？1997年，塞林格团队提出

一个方案：把两个纠缠光子对放入某种实验装置中，令光子对1中的一个光子跟光子对2中的一个光子发生纠缠（即令二者变得无法区分），二者构成新的纠缠关系；俘获这个新的纠缠光子对中的一个光子，则剩余的三个光子便会彼此纠缠。1999年，塞林格团队首次实现了三光子纠缠，在该团队工作的中国留学生潘建伟作为研究骨干，为这一工作做出了重要贡献。



潘建伟（中国科学院院士）

(https://www.sohu.com/a/191058951_703572)

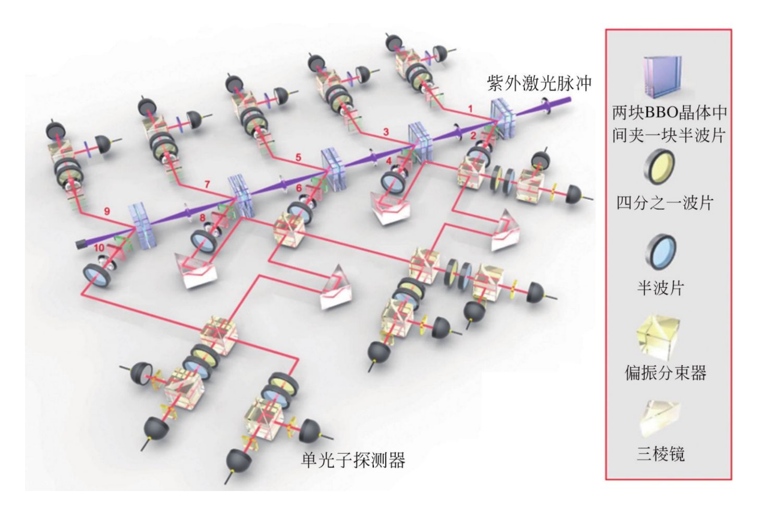



图18-7　十光子纠缠实验装置图

2000年，潘建伟学成回国，在中国科学技术大学建立了实验室，

以其为代表的中国科学家在光量子信息处理领域走到了国际的最前沿，开始不断创造多光子纠缠的纪录。从2004年开始，潘建伟团队依次实现了五光子、六光子以及八光子的纠缠。2016年，潘建伟团队再度打破纪录，成功实现十光子纠缠（图18-7）。

随着纠缠光子数的逐步增加，多光子纠缠被科学家们广泛应用到量子力学理论检验、量子计算机、量子保密通信、量子隐形传态等各个方面，极大地引领和推动了量子信息科学的发展。2022年，诺贝尔物理学奖授予了克劳瑟、阿斯派克特和塞林格三人，以表彰他们“用纠缠光子进行实验，证实贝尔不等式不成立并开创量子信息科学”。

### 扩展阅读

纠缠态能不能超光速传递信息？

面对纠缠态“幽灵般的超距作用”，人们最好奇的一个问题就是：纠缠态到底能不能超光速传递信息？

笔者认为，答案是不行。纠缠态粒子虽然可以瞬时同步改变状态，但并不能传递有效信息。因为我们虽然能通过测量让纠缠态粒子从叠加态变成确定态，但却无法控制它们变成哪一种确定态，这种测量结果是随机的，因此并不能传递有效信息。

例如，我们还用上面的孪生光子来传递信息，现代数字信息都是由“0”和“1”组成的二进制代码序列，假设我们要传递信息“10”。如图18-8所示，甲乙双方事先约定偏振片垂直放置，把光子的垂直偏振态作为“1”，水平偏振态作为“0”，然后甲制备2对孪生光子并测量自己这一侧的光子，如果甲能控制测量结果按设定的规律“10”变化，那就好办了，关键是，甲只能观察结果是“0”还是“1”，而没法控制它变成“0”或者“1”。于是，甲测量完以后，可能得到的是“11”，乙的测量结果同样也是“11”，甲只能打电话（或其他经典的信息传递方式）告诉乙，这次作废，重来。重来可能得到的又是“00”，还不对，再重来。下一次，可能终于得到了“10”，可甲还得打电话告诉乙，这次对了，有效。如果甲不打电话，乙就

不知道对错。显然，只有配合打电话（或其他经典的信息传递方式）才能传递有效信息，这样一来，纠缠态信息传递速度还是不能超过光速。




图18-8　纠缠态传递信息示意图



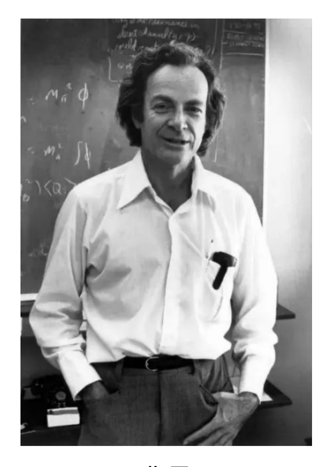
量子纠缠能超光速传递信息吗？

# 第六篇　量子·新发展

# 19 无路不走

如果给量子力学的创始人分一下代的话，普朗克、爱因斯坦和玻尔应该算是第一代量子大师，1900—1913年，他们把量子理论引入了物理学；德布罗意、海森伯、薛定谔、玻恩、狄拉克和泡利应该是第二代量子大师，1923—1930年，他们建立了量子力学的理论体系。这时候，量子力学的大厦已经基本成型了，大多数人只能添砖加瓦，但是谁也没想到，有人还能直接加盖一层楼，这个人就是第三代量子大师——理查德·费曼（1918—1988）。没错，他就是我们前面已经多次提到过的写了著名的《费曼物理学讲义》的费曼。

费曼在小学就表现出过人的数学天分，被称为“数学神童”。1935年，他进入麻省理工学院学习数学和物理，一入学就开始自学狄拉克的《量子力学原理》，书中的一句话成了他后来一生的信条，只要碰到棘手的问题，他就会习惯性地吟诵这句话：“看来这里需要全新的物理思想。”



费曼

1939年，费曼毕业后进入普林斯顿大学，师从约翰·惠勒（1911—2008）攻读研究生，选定了量子场论作为研究方向。

量子场论在1927—1928年就出现了。量子场论的奠基人不是别人，正是狄拉克。约丹、维格纳、海森伯和泡利等都做出了重要贡献。我们知道，经典的电磁场理论很好地解释了光的性质。电场和磁场的振动就是电磁波，电磁波就是光波。在量子理论诞生之后，物理学家们认识到，光子就是电磁场携带能量的最小单元，即光子是电磁场的场量子。于是他们进一步推测，既然电磁场的场量子是一个基本粒子，那么电子是不是也是某个场的场量子？很快，他们就发展了相关理论，指出电子也可以看作是电子场的场量子。进一步地，每一种基本粒子都可以看成是一种独特的场的量子化的表现形式。于是，量子场论逐渐发展起来了。

1929年，一个新的名词出现了——量子电动力学。“量子电动力学”这个名字听起来挺吓人，但研究内容并不可怕，简单来说，它是关于光和物质相互作用的量子理论。

量子电动力学诞生之初，遇到的最大困难就是在计算过程中总会出现无穷大。在量子场论中，电子被认为是没有大小的点粒子，这导致随着电子的半径趋向于零，电子的质量和电荷将会变得无穷大。狄拉克在《量子力学原理》中那句“看来这里需要全新的物理思想”，就是针对无穷大问题来说的。

费曼决定解决这个问题。大部分物理学家都认为他们面临的困难主要在于数学方面，但是，量子电动力学所需要的数学越来越艰深，深得让物理学家们望而生畏。费曼决定另辟蹊径，跳过抽象的数学，用图像化的方法来解决问题。最后，他成功地创立了“路径积分”的新方法，发明了费曼图直观地处理各种粒子的相互作用，并且提出了“重正化”的数学技巧，一举解决了这一难题，得到的计算结果与实验结果达到了惊人的一致性。

例如，有个描述电子自旋的物理常数叫g因子（一个磁矩和角动量之间的比例常数），在狄拉克理论中的数值应该是2，而费曼的计算预测g因子数值为2.002 319 304 76。目前所测的实验值是2.002 319 304 82，这个预测结果是如此惊人的准确，不由得人们不承认费曼理论的正确性。用费曼的话来说，这一精度相当于测量纽约与洛杉矶之间的距离而误差只有一根头发丝的粗细。

费曼总是能用最简洁的图像或者语言描述最复杂的物理现象，具有透过现象看本质的本领。尽管量子电动力学的理论艰深复杂，但当人们问及他关于光与电子相互作用的量子机理时，他只用了三句话就道出了其中的精髓：第一，光子从一处到另一处的行为存在着概率关系；第二，电子从一处到另一处的行为也存在着概率关系；第三，吸收电子还是发射光子同样存在着概率关系。他说，如果你能找到这些概率关系的话，你就会知道电子和光子在相互作用时该发生什么事了。

费曼曾写了一本书《QED：光和物质的奇妙理论》（QED是量子电动力学的英文缩写），书中简单介绍了他的理论。例如，在探讨一个光子从S点经镜面反射到P点的路径时，人们通常认为，光沿着直线传播，光子也应该是这样（图19-1（a））。然而，这个结论却是“错误”的。费曼指出，单个光子运行的特征是“概率性”的，在从S点到P点的运行过程中，光子的轨迹有着许许多多的可能性，或者说，有着一切路径的可能性（图19-1（b））。通过对所有路径的概率幅进行求和，就能得到光的最终概率幅，从而得出光走的是用时最短的路径的结果。这就是费曼对于光的反射的解释，也体现了费曼的路径积分的思想。费曼说：“光并不是真的只沿一条直线前进，它能‘嗅出’与之邻近的那些路径，并在行进时，占用直线周围的一个小小的空间。”

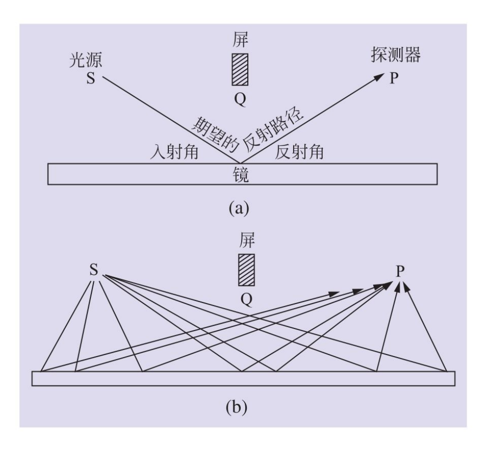



图19-1　光的反射

（a）经典光学路径；

（b）费曼的路径积分图像

经过费曼的发展，“路径积分”获得了巨大的成功，已经成为量子力学的新的数学表示形式。这样，量子力学就有了三种数学表示形式——波动力学、矩阵力学和路径积分。从数学方法上来说，矩阵力学使用矩阵，波动力学使用微分方程，路径积分则是使用积分的、整体的观念来解释和计算量子力学。并且，路径积分的方法有一个很大的优点：可以很方便地从量子力学扩展到量子场论。因此，路径积分已经成为现代量子场论的基础理论。

创立夸克模型的盖尔曼曾这样评价：“量子力学的路径积分形式比一些传统形式更为基本，因为在许多领域它都能被应用，而其他传统表达形式将不再适用。”

路径积分为什么会受到物理学家如此青睐，它的魅力到底是什么呢？答案是：它可以更形象、更直观地分析量子力学与经典力学的联系，它更能够体现物理体系的整体性质。费曼从经典力学的作用量与量子力学中的相位关系出发，把经典作用量引进到了量子力学，得出了粒子在某一时刻的运动状态，取决于它过去所有可能的历史的结论，从而给出了解决量子力学问题的新途径。其核心思想是：从一个时空点到另一个时空点的总概率幅是所有可能路径的概率幅之和，每一条路径的概率幅与该路径的经典力学作用量相对应。

作用量是一个很特别、很抽象的物理量，它表示一个物理系统内在的演化趋向，能唯一地确定这个物理系统的未来。只要设定系统的初始状态与最终状态，那么系统就会沿着作用量最小的方向演化，这被称为最小作用量原理。例如，光在从空气进入水中传播时，它所走的路径是花费时间最少的路径。

把作用量引进量子力学，费曼便架起了一座联结经典力学和量子力学的新桥梁。为了让读者更好地体会路径积分的魅力，我们仍然通过双缝实验来对其思想进行说明。在此，我们要把经典力学的路径和量子力学的概率幅结合起来分析。

以前我们在讨论双缝实验的叠加态时，只考虑了通过狭缝1的状态$\psi_1$和通过狭缝2的状态$\psi_2$的叠加，但是$\psi_1$和$\psi_2$仅仅显示了电子在双

缝处的状态，而电子从出发到双缝，以及从双缝到屏幕的过程并没有显示，也就是说，$\psi_1$和$\psi_2$是两种汇总了的状态，即使电子从出发到屏幕有千万条路径，只要通过狭缝1就被汇总到$\psi_1$中，只要通过狭缝2就被汇总到$\psi_2$中。

费曼对此展开了追问，如果我们观察电子从出发到屏幕的全过程，会是什么图景？

如图19-2所示，电子枪发射一个电子。在经典运动方式下，电子从$A$出发落到屏幕上任意一点$B$时只能通过1、2两条路径到达，而根据电子的量子特性，电子在$B$点出现的概率幅$\psi$是路径1的概率幅$\psi_1$和路径2的概率幅$\psi_2$之和：

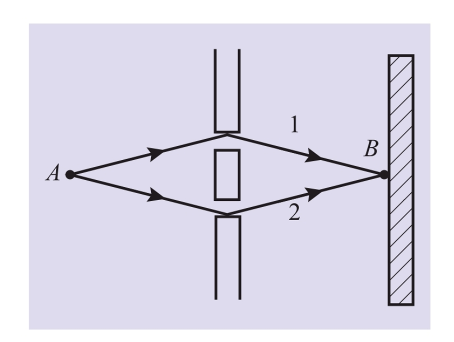



图19-2　按经典运动考虑，电子有两条可能路径；按量子特性考虑，落点概率幅是两条路径的概率幅叠加

$$\psi=\psi_1+\psi_2$$

下面来设计一个稍微复杂一点的情况，在双缝和屏幕间再插入一块板，板上有三条狭缝，如图19-3所示。按经典路径，那么现在从$A$到$B$有6条可能路径。于是电子在$B$点出现的概率幅就是从路径1到路径6的概率幅之和：

$$\psi=\psi_1+\psi_2+\psi_3+\psi_4+\psi_5+\psi_6$$

现在，让我们想象一下，如果在插入的板上刻出更多的狭缝，4条、5条、6条……两条狭缝之间的距离越来越小，当狭缝的数目趋于无穷时，会有什么效果呢？对了，那就是——这块板不见了，就跟没有这块板一样！

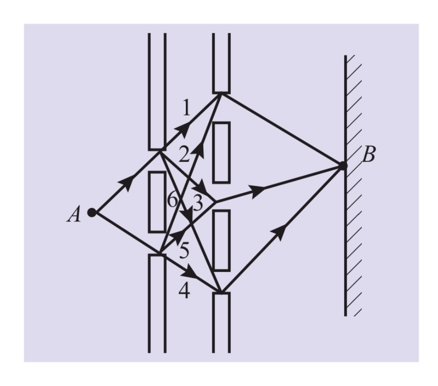



图19-3　双缝和屏幕间插入一块刻有三条狭缝的板，电子有6条可能路径

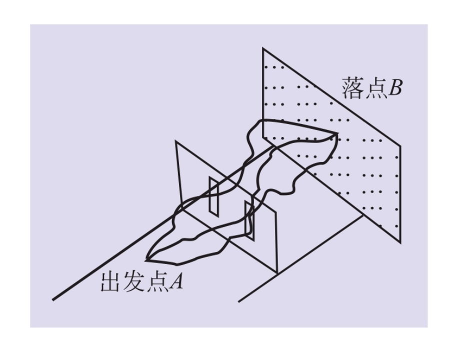



图19-4　电子路径是无数种可能路径的叠加

虽然空空如也，但我们可以认为在从$A$到$B$的空间里插满这种有无穷条狭缝的板，那么电子就在这些板之间来回碰撞转折，于是有无数条可能的路径实现从$A$到$B$的过程，如图19-4中给出的3条可能路径。所以，在双缝干涉实验中，电子在$B$点出现的概率幅就是空间中所有可能路径的概率幅之和：

$$\psi=\psi_1+\psi_2+\psi_3+\ldots$$

我们知道，积分运算正是处理这种问题的好方法。费曼通过他的路径积分计算表明，当把所有可能路径都考虑进去时，算出的概率跟实验值刚好吻合。

这就是路径积分理论对于双缝实验的解释，也就是说，电子最终的落点是由所有可能路径决定的，因此，即使只发射一个电子，它也会落到双缝干涉位置上去。

需要注意的是，电子有无数条可能的路径，但它并不是选择其中的一条，而是无数条的叠加，这是态叠加原理的体现，显然，叠加后它没有明确的运动轨迹，这也是不确定原理的必然结果。

1942年，费曼完成了博士论文，这篇论文初步提出了路径积分方法，他的导师惠勒对此大为赞叹。因为爱因斯坦也在普林斯顿，所以惠勒将费曼的论文拿去给爱因斯坦看。他对爱因斯坦说："这论文太精彩了，是不是？你现在该相信量子论了吧？"

爱因斯坦看了论文，沉思了一会儿，说："我还是不相信上帝会掷骰子……可也许我现在终于可以说是我错了。"



双缝实验的路径积分解释

# 20 平行世界

根据薛定谔方程演化的量子态，并不会自然地出现波函数坍缩这样的现象，因此，波函数坍缩实际上是独立于量子力学基本框架之外的一个额外假设，这也是它引起争议的主要原因。事实上，波函数坍缩的主要问题出在“突变”上，为了消除这个破绽，科学家们各显神通。第14章所述的退相干理论就为波函数坍缩找到了一个合理的演化过程，从而不需要“突变”，这一理论也得到了普遍的接受。但是，这并不是目前唯一的理论，在退相干理论出现之前，已经出现了一个神秘的理论——多世界理论。

1953年，费曼的导师惠勒招收了一个新的博士生休·艾弗雷特（1930—1982），他是从数学系转过来的，而他的本科学位是化学工程，这样，数理化样样不落的艾弗雷特就成了比费曼低十几届的师弟。

艾弗雷特从小就喜欢读科幻小说，喜欢琢磨一些古怪的问题。他12岁时曾给爱因斯坦写信，声称自己解决了一个难题——当不可抗拒的力碰到不可移动的物体时会发生什么（这个问题有点类似于最强的矛攻击最强的盾会发生什么）。爱因斯坦觉得这孩子很有趣，竟然给他回信了。爱因斯坦在信中写道，世界上虽然没有什么不可抗拒的力和不可移动的物体，但却有一个固执的小男孩，他故意为自己制造了一个奇怪的难题，然后费力地走上了解决它的道路。

艾弗雷特进入普林斯顿后，他最开始的兴趣在博弈论方面，并且在1953年发表了一篇关于博弈论的论文。1954年秋天，玻尔访问了普林斯顿，玻尔是惠勒在欧洲留学时的导师，所以艾弗雷特有幸和玻尔近距离接触，了解了量子力学的测量难题。通过进一步和周围同学以及玻尔的助手讨论，艾弗雷特觉得波函数坍缩实在令人难以接受，于是，他决定将量子测量作为自己的博士研究课题，另起炉灶。



艾弗雷特

作为一个科幻爱好者，艾弗雷特从小就天马行空的古怪思维发挥作用了，不到半年，他就想到了一个好点子，这是一个前所未有的新思想——他很干脆地直接否定了波函数坍缩，提出宇宙诞生之初就产生了宇宙波函数，而且宇宙波函数会持续演化下去，根本不会发生波函数坍缩，只会发生不断的分裂，变成越来越复杂的叠加态。

当艾弗雷特第一时间把自己的想法告诉惠勒时，惠勒大吃一惊，他非常反对艾弗雷特用“分裂”一词来描述世界的一分为二。但是，他并没有阻止艾弗雷特继续研究下去。惠勒对于教育有特殊的理解。“大学里为什么要有学生？”惠勒说，“那是因为老师有不懂的东西，需要学生来帮助解答。”所以他并不过多干涉学生的研究自由。

1957年，艾弗雷特终于完成了博士论文，他把自己的理论叫做“普适波函数理论”。他的理论是，所有孤立系统的演化都遵循薛定谔方程，但波函数坍缩从不发生。整个宇宙的波函数是由一系列平行世界波函数叠加而成，这些平行世界各自独立演化互不干扰（图20-1）。因此，后来人们把他的理论改称为“多世界理论”。

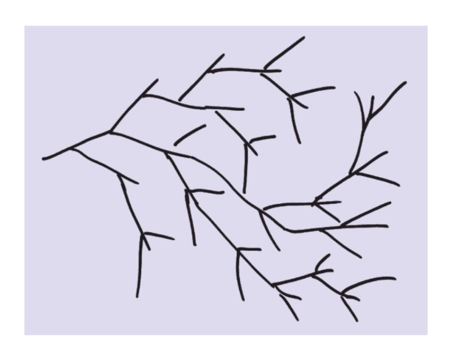


图20-1　平行世界分裂示意图

按照艾弗雷特的理论，自然界不再有量子和经典的区分，宇宙中的所有物体无论大小都由波函数描述，所有物体都处于叠加态。在他看来，被测系统、测量仪器和观察者都有自己的波函数，也都存在各种状态，于是这三者构成的整体也就存在各种叠加态，这些叠加态中每个状态都包含一个确定的观察者态、一个具有确定读数的测量仪器态，以及一个确定的被测系统态，因此，在每一个状态中的观察者都会看到一个确定的测量结果，这样，在这个状态中的测量者以为发生了波函数坍缩，其实是因为他们不知道其他平行状态的存在而已。实际上从整体来看，波函数并没有坍缩，它仍然在各种平行状态中发展着。

按照多世界理论，“薛定谔的猫”处于两种世界的叠加态：一种世界里猫是活的，另一种世界里猫是死的。这两种世界一样真实，并行存在，而且这两种世界会独立演化，互不影响。在一种世界里，当观察者打开箱子，他会看到一只活猫；在另一种世界里，当观察者打开箱子，他会看到一只死猫。在两个世界里的观察者都以为波函数发生了坍缩，是因为他们都感觉不到另一个世界的存在。

惠勒虽然觉得艾弗雷特的论文难以理解，但他采取了包容的态度，并把艾弗雷特的论文寄给玻尔审阅，结果，遭到了玻尔的激烈反对。其他科学家的态度也和玻尔差不多，甚至有人嘲讽其为“彻头彻尾的精神分裂症”。不出所料，艾弗雷特的论文发表以后，受到了学术界的冷遇，没几个人对此表示关注。

艾弗雷特毕业后，惠勒邀请他留校任教，但艾弗雷特拒绝了，因

为他发现自己并不喜欢搞学术研究。艾弗雷特在美国国防部找了一份工作，从此离开了物理领域，再也没有发表过一篇论文。正像徐志摩一首诗中说的那样："悄悄的我走了，正如我悄悄的来；我挥一挥衣袖，不带走一片云彩。"

令人惊讶的是，到了20世纪70年代，一度备受冷落的多世界理论竟然又复活了，在支持者的宣传下，这一颇具科幻色彩的理论迅速走红，不但获得了大量科幻迷的追捧，连理论物理学家中也有不少人成为这个理论的拥趸，当然，反对者也并不少。

在多世界理论中，宇宙从来不会做选择，它只是按照概率不停地分裂为更多的世界，这样，从中很容易就会推出一个怪论：一个人永远不会死去！在世界的不断分裂中，人总在某个分支世界中活着，这个怪论被美其名曰称为"量子永生"。当然，"薛定谔的猫"也永远会在某一个宇宙分支里活着（图20-2）。

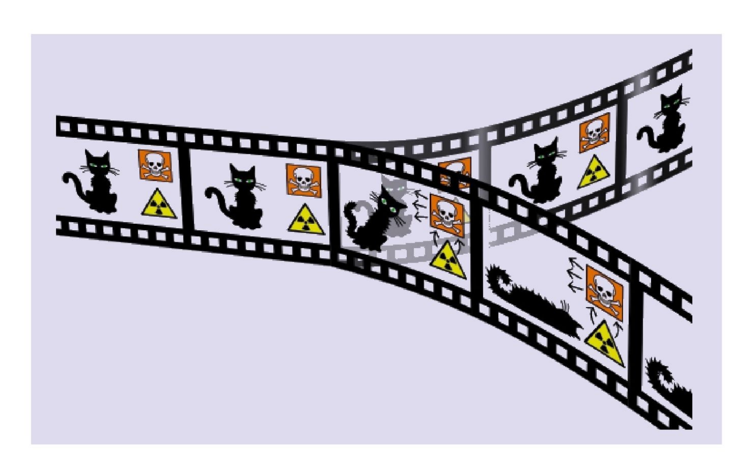


图20-2　量子永生示意图

（https://www.sohu.com/a/312171762_99994982）

"量子永生"使多世界理论看上去似乎很美好，谁不想永远活下去呢。然而，有一个问题却使多世界信奉者苦恼：为什么我们感觉不到平行世界？没有任何人能证明平行世界的存在，那么，它到底是一种数学技巧还是物理实在呢？

笔者认为，波函数并非物理实在，即使宇宙处于多种状态的叠

加，也只不过是波函数的叠加，这并不能看作是多个宇宙的实体存在，所以平时世界只是存在于数学里，并不是存在于物理里。反过来，假如多世界是物理实在的话，那么当世界一分为二时，诞生一个新世界的能量从何而来？总之，这并不是一个让人容易接受的理论。有物理学家评价说，多世界的假设很廉价，但宇宙付出的代价却太昂贵。

艾弗雷特有两个孩子，老大是女儿，老二是儿子。艾弗雷特和妻子在子女教育上观念非常一致：孩子们应该不受任何管束，自由成长。结果，他的女儿成了问题女孩，养成了吸毒等恶习，让他追悔莫及。1982年，不满52岁的艾弗雷特死于心脏病突发。1996年，他的女儿自杀了，她在遗书中写道，她希望能在另外一个平行世界里与父亲相会。他的儿子2007年接受BBC采访时表示："父亲不曾跟我说过有关他的理论的片言只语……他只活在自己的平行世界中。"

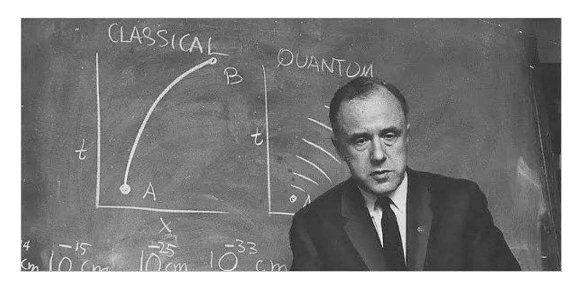


双缝实验的多世界解释

# 21 历史能改变吗

回顾量子发展史，有一个很有趣的现象，自从爱因斯坦来到美国普林斯顿以后，这里就成了量子力学新思想的发源地。爱因斯坦自己提出了量子纠缠；玻姆提出了新的隐变量理论；费曼提出了路径积分理论；艾弗雷特提出了多世界理论。玻姆可以说是受到了爱因斯坦很深的影响，而另外两人，费曼和艾弗雷特，就不能归功于爱因斯坦了。他们的脱颖而出，要归功于他们的导师——惠勒。

费曼和艾弗雷特作为惠勒的得意门生，在公众中的知名度都很高，这说明惠勒是一个优秀的导师，在培养人才方面是首屈一指的。但惠勒本人却没有那么高的知名度。很多人都听说过“黑洞”“虫洞”和“量子泡沫”等词汇，但是，很少有人知道，这些词汇都是惠勒发明的，这些词汇原来都有一个佶屈聱牙的专业术语，例如，如果有人提到“引力坍缩星体”和“爱因斯坦-罗森桥”，你还会对此感兴趣吗？而这正是“黑洞”和“虫洞”原来的名字。正是惠勒发明了这些通俗形象的词汇以后，这些词汇才得以走红世界，被大众所熟知，对激发公众的科学热情起到了极大的推动作用。




惠勒

惠勒一生的研究范围非常广泛，涉及核物理、核武器的设计、广义相对论、相对论天体物理、量子力学、量子引力及量子信息等领域。他曾经在哥本哈根跟随玻尔从事博士后研究，与玻尔一起发展出原子核分裂的“液滴模型”，并用它发展了核裂变理论；他也参加过“曼

哈顿计划”，是最早研究原子弹的美国人之一。在他开始研究广义相对论以后，提出了一句广为人知的话来概括广义相对论：“物质告诉时空如何弯曲，时空告诉物质如何运动。”这句话简洁形象地表达了广义相对论的核心思想，让普通人也能一下子就明白。

1979年3月14日，普林斯顿大学召开了纪念爱因斯坦诞辰100周年的学术讨论会，在这次会议上，惠勒针对量子测量问题提出了一个惊人的实验构想——延迟选择实验。

延迟选择实验就是说，先不固定实验设置，等快要测得实验结果的时候再决定实验设置。在一般的测量实验中，实验设置都是提前固定好的，这样所有路径的可能性其实已经提前预设好了，但惠勒想“延迟”这些路径的出现。举例来说，做双缝实验，先打开一条狭缝，等电子通过以后再打开另一条狭缝，然后再观察电子在屏幕上的落点。惠勒的问题是，这时候还会出现干涉图样吗？也就是说，惠勒要“延迟”电子的选择，迫使电子在通过狭缝以后再来选择是通过一条狭缝还是通过两条狭缝。

这个想法太疯狂了，立即引起了学术界的兴趣。随后几十年中，他的思想实验变成了现实，物理学家们利用光子成功进行了多种延迟选择实验。其中有一个近乎理想化的延迟选择实验，也被称为量子擦除实验，其实验结果令人震惊。

光学实验中，常用到一个光学器件叫分束器，它能使入射到它上面的光一半透射一半反射（图21-1）。有一种特殊的分束器叫偏振分束器，它可以按光的偏振态分束，使透射光和反射光全部变为偏振光，且两束光的偏振方向相互垂直（图21-2）。

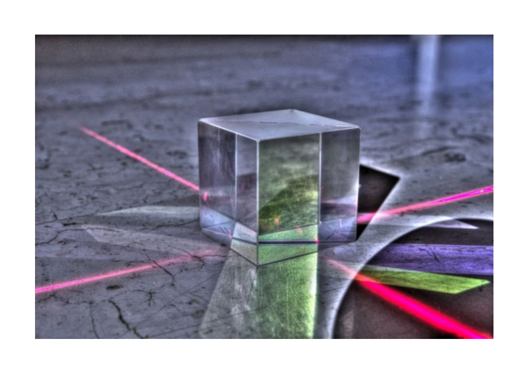


图21-1　分束器

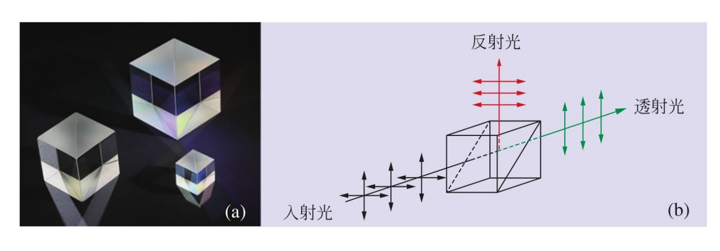



图21-2　偏振分束器和偏振分束器分光原理图

（a）偏振分束器；（b）分光原理

(http://www.shanghaioptics.com/product/29.html)

量子擦除实验中，就要用到偏振分束器，而且是发射一个一个的单个光子射向分束器。一束光被分为透射和反射两部分很好理解，但是如果是一个光子射向分束器，它会如何前进呢？如果按照经典想法，我们可能觉得它只能选择透射路径或者反射路径其中之一，但是根据量子力学原理，我们知道，如果你不去测量，它应该处于透射路径和反射路径的叠加态（尽管这很难理解，但实验事实就是如此）。

量子擦除实验简化示意图见图21-3，光子经过偏振分束器B1，处于路径1和路径2的叠加态，经反射镜反射后，两条光路在另一个偏振分束器B2处汇聚，由于两条光路偏振信息不同，具有可区分性，所以不会发生干涉，这样探测器D1和D2就可以随机探测到通过路径1和路径2的光子，如图21-4（a）所示。

事实上，如果这两条光路不携带偏振信息的话，两条路径的光子就没有可区分性，在B2处交汇以后就会发生干涉。于是，这个实验的重点来了，科学家们在探测器D1和D2前放置了两台消偏器，只要打开消偏器，两条路径的偏振可区分性就会被消除，重新变得不可区分，结果令人大吃一惊，探测器上出现了干涉图样，如图21-4（b）所示。

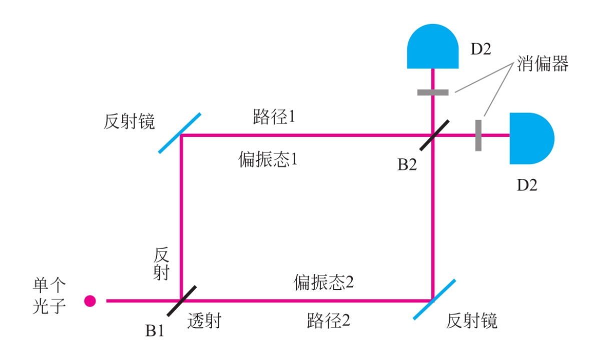



图21-3　量子擦除实验简化示意图

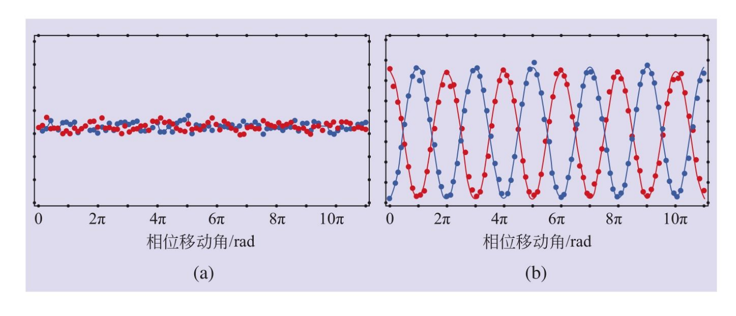



图21-4　实验结果

（a）不干涉；（b）干涉

要知道，消偏器是在B2的后边，光子如果要干涉只能借助B2来实现，现在它已经过了B2的位置，按我们的日常经验，即使这时候消除偏振，也应该无法干涉了。但是光子居然因为加了个消偏器而继续干涉，这实在是太不可思议了。

我们该如何理解这个实验现象呢？费曼曾经说过："我想我可以相当有把握地说，没有人能理解量子力学。"不过，如果用他的路径积分理论来分析，似乎可以强行"理解"光子的表现。路径积分理论指出，粒子在某一时刻的运动状态，取决于它过去所有可能的历史。那么，光子通过消偏器以后，它所有可能的历史都发生了变化，最后的测量结果自然就会表现为干涉。正因为如此，路径积分也被称为"历史求和"。但是，如果进一步追问，光子之前走过的路程算不算历史？它是如何被改变的？我们该如何回答呢？我想恐怕现在科学家们还拿不出一个令人满意的答案。

延迟选择实验彻底冲击了人们关于“实在”或“真实”的传统观念，它使人们看到“观察”能改变所谓的“实在的过去”！这彻底改变了人们对“历史”的看法，所谓“客观实在性”在这一实验面前被动摇。这时候，玻尔常说的一句话似乎让人有了更深的体会：“物理学不能告诉我们世界是什么，我们只能说，观察到的这个世界是什么。”

惠勒晚年一直在思考这些“本源”问题。他说：“我无法阻止自己去思考‘存在’之谜，从我们称之为‘科学’的理论推演与实验，到这个最宏大的哲学命题，链条一环扣一环，在探索整个链条的道路上，并不存在特殊的一环，能叫一个真正有好奇心的物理学家说，‘我就到这儿了，不往前走了。’”

# 第七篇　量子·幕后英雄

# 22 洞悉固体

量子力学的建立，从根本上改变了人们对物质结构的认识，使许多物理现象得到了明确的解释。从此，量子力学开始在现代高科技领域发挥重要作用，例如，通过固体物理的量子理论，人们明白了半导体的原理，而对半导体的研究又导致了晶体管和芯片的发明，从而为现代电子信息工业的发展奠定了基础。又例如，人们在量子力学的帮助下解释了物质磁性的来源，从而发展出了磁存储技术，于是发明了电脑的机械硬盘。再如，激光器也是根据光的量子辐射理论而发明的。

上面这些例子中，我们很少有人意识到这是量子技术，因为量子力学只是这些器件的幕后英雄，这些技术实现的功能里并没有体现出量子特征，这些器件遵从经典物理的运行规律，属于源于量子力学的经典技术。

固体物理是现代高技术科学（如半导体电子学、激光物理、材料科学等）的重要基础，如果没有固体物理的理论指导，人类可能很难步入现代这样一个由大规模集成电路主导的信息社会，而现代固体理论的发展完全得益于量子力学的应用。

爱因斯坦是将量子理论引入固体物理中的第一人。1907年，爱因斯坦利用能量量子化解释了固体比热问题。人们早就发现，固体的比热会随着温度的降低而大幅度减小，但用经典物理却完全无法解释这个现象。爱因斯坦意识到，固体中原子振动的能量也是一份一份的，是量子化的，从而很好地解释了这个问题。

固体量子理论的研究对象主要是晶体，因为晶体内部原子排列有序，规律性强。根据周期性排列的最小单元，可以将晶体看作是一系列相同晶格的重叠堆积，如图22-1所示。

晶体里的原子并不是静止不动的，它们不停地在各自的平衡位置附近做微小的振动，由于晶体中原子间有着很强的相互作用，因此，

一个原子的振动会牵连着相邻原子随之振动。如果把原子比作小球的话，整个晶体犹如许多小球在三维空间中规则排列，而小球之间又彼此被弹簧连接起来一样（图22-2），因此，每个原子的振动都要牵动周围原子振动，使振动以弹性波的形式在晶体进行中传播，这种波被称为晶格振动波，简称“格波”。

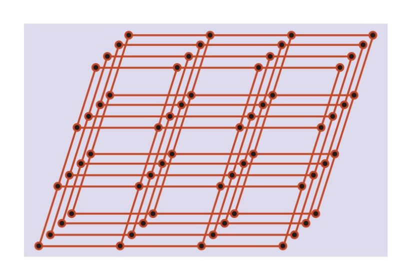


图22-1　晶体由晶格并置堆积而成（图中的平行六面体就是晶格）

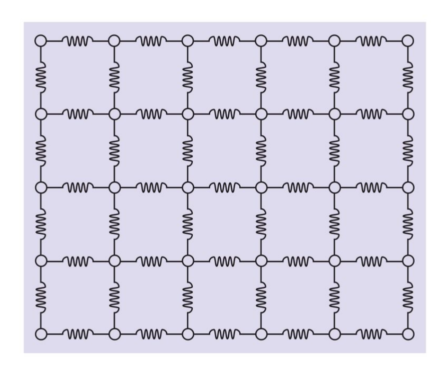


图22-2　晶体中原子与原子之间就像用弹簧连着一样

爱因斯坦假定，原子振动可以看作是一种简谐振动，所有原子都具有相同的振动频率。

经典的简谐振动我们很熟悉，把一个小球系在弹簧上，把它拉开平衡位置以后松手，小球来回往复运动，这时候，按经典力学，体系的能量$E=\frac{1}{2}kA^2$（$k$为弹性系数，$A$为振幅），如果把不同振幅下的能量画成曲线，为一条抛物线，随振幅不同，能量可以从0开始连续变化，如图22-3（a）所示。

但是，在原子的简谐振动中，振动能量却是量子化的，图

22-3（b）给出了一个振动的原子的能级分布图。可以看到，能量只能取图中$E_0$、$E_1$、$E_2$、$E_3$等一系列分离的能级，而且最低能级$E_0$不为零。图中还给出了每个能级对应的波函数平方的图像，显示了原子在不同位置上出现的概率密度，可见其运动特征和经典的弹簧振子是完全不同的。

格波是晶体中全体原子都参与的集体振动，既然单个原子的振动能量是量子化的，那么格波的能量自然也是量子化的。

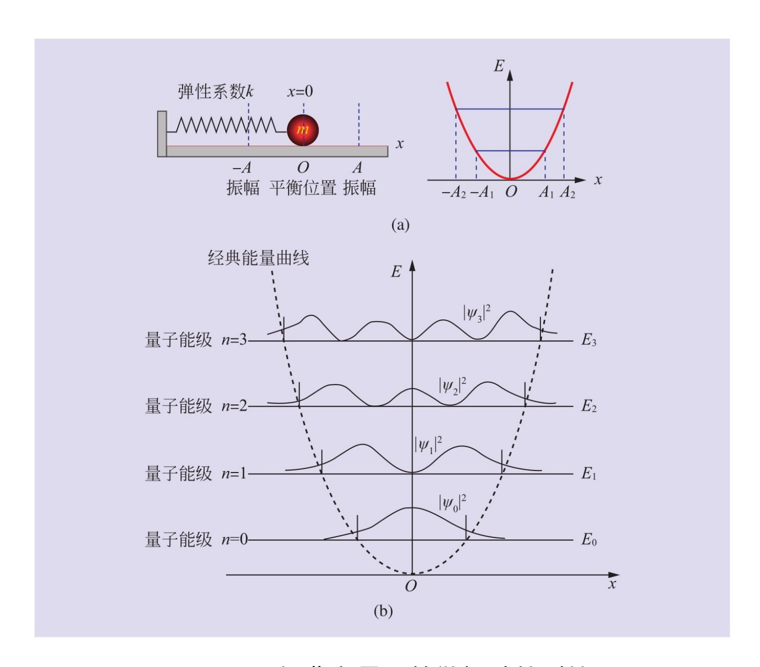



图22-3　经典和量子简谐振动的对比

（a）弹簧谐振子及其能量曲线；（b）量子谐振子的能级和概率密度分布图

1930年，苏联物理学家塔姆（1895—1971）在研究格波时，突然想到了波粒二象性。他想，既然像电子这些原本只能用粒子来描述的东西也能用波描述，那么原本只能用波描述的东西是不是也可以用粒子来描述呢？于是，他就设想把格波的最小能量单位与一种假想的粒子对应起来，称之为“声量子”，后来人们改称为“声子”。

声子是将波动量子化的粒子，它并不是像光子和电子那样是“真实”的粒子，而是一种人为假设的准粒子。但是声子却似乎具有“真

实”的量子粒子的一些属性，将晶格振动看作是声子的运动，可以很好地解释固体物理中的很多现象。例如，格波间的相互作用可以看作是声子间的碰撞。再例如，当研究电子与晶格的相互作用时，若电子从晶格获得能量，可看作是吸收声子；若电子给予晶格能量，可看作是发射声子，这样处理问题就方便多了。再如，可以把固体看作是包含有“声子气体”的容器，从而可将气体分子运动论和量子统计力学的处理方法用于处理固体问题。另外，声子在超导现象的解释中也扮演了关键角色。

总之，声子这个概念出现以后，极大地推动了固体物理的发展，它现在已经成为固体物理学的基本概念。

从爱因斯坦提出光波具有波粒二象性，到德布罗意提出实物粒子具有波粒二象性，再到塔姆提出格波具有波粒二象性，波和粒子似乎总是相伴相生。回顾这段历史，也许能让我们对波粒二象性的物理内涵有更深刻的认识。

用量子理论来研究固体的另一条主线是能带理论的发展。1926年，薛定谔提出薛定谔方程以后，化学家们立刻开始用它来计算分子中电子的运动并在几年之内就发展出了一系列化学键理论。20世纪30年代，科学家们开始用薛定谔方程计算晶体中电子的运动，固体的能带理论随之建立起来。

固体是由大量微观粒子组成的复杂体系，原子数达到$10^{23}$的数量级，电子数目更是庞大，而科学家们就是要通过如此庞大体系的微观粒子的运动规律阐明固体的宏观物理性质。这个体系是非常复杂的，大量电子之间会相互影响、相互作用，但是其基本特点不会变，那就是每个电子都在一个具有晶格周期性的势能场中运动。于是，通过一系列简化与近似，薛定谔方程就可以近似求解了。

读者还记得，求解单个原子的薛定谔方程得到的是一系列分离的能级，而晶体中得到的则是一系列分离的能带，这些能带是由大量原子能级叠加组合而成的，由于能级间隔非常小而可以看作是连续的能带（图22-4）。这些能带与整个晶体而不是单个原子联系在一起，于是，如果一个能带没有被电子全部占满，电子就可以在电压作用下在

整个晶格中到处移动，这个晶体就能导电。如图22-4所示，金属的最高能带没有被电子占满，所以它们是良导体。半导体和绝缘体的能带都是要么被电子占满（满带），要么没有电子（空带），所以没法导电，但半导体的满带和空带之间的能带间隔较窄，所以电子可在热或光的激发下从满带跃入空带，使原来的满带和空带都成为不满的能带而导电（图22-5）。

人们发现，在纯的半导体材料（本征半导体）中掺入某些杂质，可以极大地提高其导电能力，利用这种特性可以制成掺杂半导体，并由此制成了二极管、三极管等重要的半导体器件。

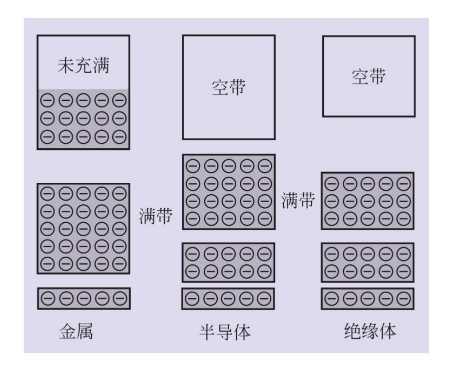



图22-4　金属、半导体和绝缘体的能带结构特征

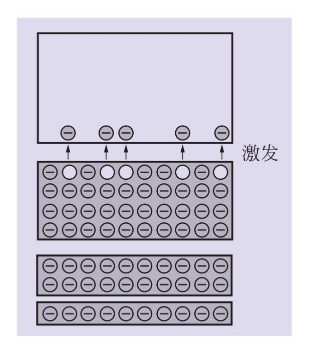



图22-5　半导体导电机理

## 扩展阅读

半导体的满带和空带之间的能带间隔较窄，电子可在热或光的激发下从满带跃入空带，使原来的空带出现少量自由电子而导电，使原来的满带出现同等数量的空穴而导电。

当满带上的部分电子被激发到空带后，留下了空穴。如果施加电压，在外电场作用下，空穴附近的电子能够移动到这个空穴中，从而在原位置留下一个新的空穴，整个近满带中大量电子的缓慢移动，就像空穴在反方向缓慢移动一样。因此，可以将空穴假想成一种带正电的粒子（只是一种准粒子），这样，近满带的导电问题就转化为少量空穴的移动导电，与导带中少量自由电子的导电问题十分相似，研究起来更为方便。

电子和空穴都能导电，为了区分方便，将它们称为N型和P型载流子（N和P分别代表英文单词Negative和Positive的首字母）。在半导体中掺入富电子或缺电子的杂质，会在能带间隙中引入额外的能级，导致N型和P型载流子数目的改变，从而形成“N型半导体”或“P型半导体”。额外能级的引入，相当于缩小了能带间隙，因此只要有少量杂质掺入，就会明显地提高半导体的电导率。例如，10万个硅原子中掺入1个杂质原子就能使硅的电导率增加1000倍左右。

能带理论让人们搞清楚了半导体的导电机理，带动相关研究快速发展起来。随着对半导体特性研究的深入，1947年，半导体材料迎来了一个重大发明——晶体管。晶体管既可以用来做电信号的放大，也可以用作电压控制的开关，由这些开关组成的逻辑电路网络可以控制电子设备或处理计算机中的信息，由此启动了电子器件小型化的进程。发展到今天，一块小小的芯片上可以集成上百亿个晶体管，如此超大规模的集成电路，使人类社会进入了信息时代的黄金时期。在我们享用便捷的手机、电脑和各种家用电器时，不要忘了，这正是量子理论带给我们的快乐。

# 23 隧穿

美籍俄裔科学家乔治·伽莫夫（1904—1968）因为提出宇宙大爆炸理论而为人们所熟知，他写的科普作品《从一到无穷大》直到现在都是畅销书，但很多人不知道，他是世界上第一个发现神奇的“量子隧道效应”的人。



伽莫夫

伽莫夫毕业于列宁格勒大学，1928年夏天，他获得了一份奖学金，到德国哥廷根大学访学3个月。当时的哥廷根正是量子力学的发源地之一，来哥廷根后，伽莫夫很快就熟悉了刚刚建立的量子力学。有一天，他在图书馆读到一篇卢瑟福写的有关原子核$\alpha$衰变（某些放射性元素的原子核释放出$\alpha$粒子的现象，$\alpha$粒子是氦原子核）的文章，他立刻意识到卢瑟福试图用经典理论来解释$\alpha$衰变的思路是完全行不通的，不可能得出合理的解释，对于这种微观粒子，必须用量子力学来处理。他立刻投入计算，没几天就写成了一篇论文，对$\alpha$衰变做了全新的量子力学分析。在此文中，他发现了量子隧道效应。

伽莫夫发现，根据量子力学的规律，即使微观粒子的能量并不足

以越过能量势垒，也会有一定的概率穿过势垒，而且粒子穿越势垒的概率可以通过薛定谔方程精确计算出来。例如，对于铀-238原子核，它放出的$\alpha$粒子的能量只有4.2 MeV，但是它有一定概率穿越能量高达35 MeV的库仑势垒从原子核里逃逸出来，这在经典物理里是绝对不可能的，就像一个人只能跳4.2 m却跳过了35 m高的墙一样，实在是太匪夷所思了。

下面我们通过一维方势垒粒子的运动来简单介绍一下量子隧道效应。

假设一个质量为$m$、能量为$E$的粒子，沿$x$轴在一维方向上运动，它受到如图23-1所示的高势能区域的阻挡（图中纵坐标表示势能，由于这个图像一堵墙，所以被称为势垒），势垒区域（0~$l$范围）的势能是$V_0$，其他区域势能为零。假设粒子从$x$轴左侧入射且$E < V_0$，那么按照经典力学，它只能在$x < 0$的区域内运动，绝无可能出现在$x > l$的区域，因为它的能量小于$V_0$，所以不可能穿过势垒，就像一颗塑料子弹不可能打穿钢板一样。

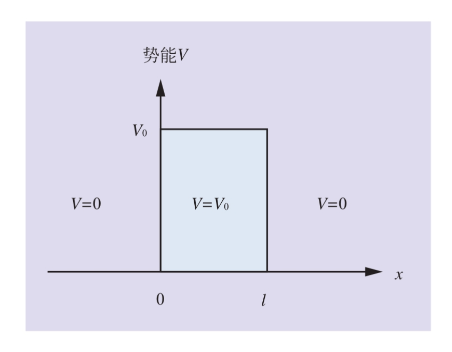



图23-1　一维方势垒

但是，把这个粒子的薛定谔方程写出来并且求解以后，结果却令人大吃一惊。图23-2给出了波函数ψ($x$)的图像。可以看到，粒子从$x$轴左侧入射，但它的波函数出现在了$x > l$的区域，表明粒子有穿透势垒的概率，这与经典力学是完全不同的。另外，波函数在势垒内部（0~$l$范围）是呈指数衰减的，这就意味着如果势垒宽度$l$变厚，粒子穿透的概率就会迅速下降。此外，粒子质量越小、粒子与势垒的能量

差越小，粒子穿透势垒的概率就越大。显然，对于宏观物体，由于它的质量太大，所以它穿透势垒的概率接近于零，量子力学与经典力学的结论趋于一致。

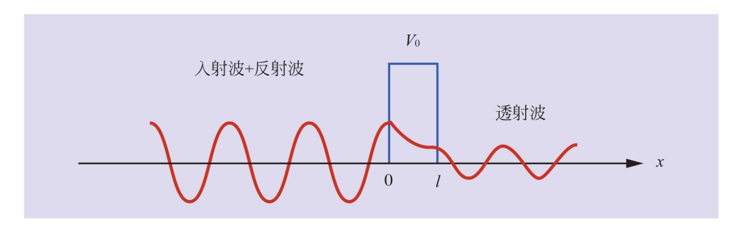



图23-2　一维方势垒粒子的波函数图像

这样，我们就能得出结论：如果微观粒子遇到一个能量势垒，即使粒子的能量小于势垒高度，它也有一定的概率穿越势垒，因为它就像是从隧道中钻出来的，所以被形象地称作隧道效应。隧道效应是一种很常见的量子效应，崂山道士的故事在量子世界里是很平常的，一点儿都不稀奇。

3个月的访学时间一晃就到了，伽莫夫不得不踏上归途。不过他没有直接回苏联，而是绕道丹麦去哥本哈根拜见玻尔。他把自己写好的用隧道效应解释$\alpha$衰变的论文拿给玻尔看，立即引起玻尔极大的兴趣。玻尔当即决定，把伽莫夫留在哥本哈根工作一年，并给了他一年的奖学金。一年之后，玻尔又介绍伽莫夫去卢瑟福的卡文迪什实验室工作。

当时，卢瑟福手下的沃尔顿和柯克罗夫特正在进行人工加速质子轰击原子核的研究。为了获得高能质子，需要通过超高电压使质子加速。他们研制了一种电压倍增电路，最高可以产生50万V的电压，但是，这已经是他们的极限了，电压再也无法升高。令他们绝望的是，根据当时的理论计算，要想使质子射入被轰击的原子核内，至少要400万V的高压，差距太大，他们已经准备放弃了。

这时候，恰好伽莫夫来了，他了解到两人的困境以后，立刻想到了隧道效应。经过几天的计算，他自信满满地告诉两人，按照隧道效应，50万V加速的质子就够用了，完全可以完成他们的实验。

两人半信半疑，不敢相信，因为从来没有听说过隧道效应，这只是伽莫夫个人提出的理论，而且看来还那么离奇。

最终，还是卢瑟福拍板，相信伽莫夫，他给沃尔顿和柯克罗夫特拨款1000英镑，让他们建造加速器，验证伽莫夫的理论设想。1000英镑在那时候并不是一笔小数目，大约相当于现在的10万美元，可见卢瑟福的魄力。1931年，伽莫夫护照到期了，只好离开卡文迪什回到苏联。1932年，沃尔顿和柯克罗夫特终于造出了质子加速器，加速器的放电管里每秒钟可以产生500万亿个质子，质子从放电管的顶部产生后，被50万V的高压加速，轰击放在放电管底部的靶子。一切如伽莫夫所料，有部分质子利用“隧道效应”穿过了原子核表面的“屏障”，进入原子核内部并引起核裂变反应。沃尔顿和柯克罗夫特的实验顺利完成，他们还验证了伽莫夫关于入射质子进入核内的概率的估算，量子隧道效应得到了有力的实验证明，从此，隧道效应被人们正式承认。1951年，沃尔顿和柯克罗夫特因上述实验获得了诺贝尔物理学奖。

隧道效应被证实以后，人们终于揭开了太阳发光之谜。我们知道，太阳发光是利用了核聚变反应。核聚变就是轻原子核聚合成稍重一点的原子核（如氢原子核聚变成氦原子核）。但是聚变反应并不是那么容易发生的，两个原子核靠近时，库仑斥力非常巨大，为了使原子核克服库仑斥力相互碰撞，需要极高的温度使原子电离并使原子核剧烈运动，以增加碰撞的概率。但是，人们发现，太阳内部的温度并没有想象的那么高，还不足以使氢原子核获得足够的动能来抵抗库仑斥力发生聚变。对此经典物理是没法解释的，所以太阳内部如何进行核聚变一直是一个谜团。而隧道效应被发现以后，则完美地解释了这一现象。库仑斥力的作用相当于一个高势垒，氢原子核即使没有足够的动能，也有穿过势垒发生聚变的概率。尽管穿越势垒的概率很低，但是太阳里的原子数量庞大，“少量”穿越势垒的粒子，已经足以使太阳发出万丈光芒。

现在，隧道效应已经成为许多物理器件的核心，如隧道二极管、约瑟夫森结、扫描隧道显微镜等。扫描隧道显微镜放大倍数可达上亿倍，分辨率达0.01 nm，它使人类第一次真实地“看见”了单个原子，

是20世纪80年代世界重大科技成就之一。

# 24 量子之眼

人类认识自然的主要信息来自于眼睛，但是，人类眼睛的分辨距离只有0.1 mm，小于0.1 mm的物体，人眼就看不清了。随着科学发展的需要，为了能够看到物质结构更小的细节，科学家们发明了显微镜。

1665年，英国的罗伯特·胡克发明了第一台光学显微镜。他用自己研制的光学显微镜观察了软木薄片，看到了木栓组织，发现它们由许多规则的小室组成，他把观察到的图像画了下来，并把小室命名为细胞，沿用至今。显微镜的发明，使人类对微观世界的认识前进了一大步。借助显微镜，人眼看到了细胞、细菌，随着对这些领域的研究，出现了细胞学和微生物学等重要学科。

1873年，德国的亥姆霍兹从理论上证明了显微镜的分辨距离与照射光的波长成正比。光学显微镜所用的光源是可见光，其波长最小是400 nm，所以光学显微镜的极限分辨距离是几百纳米，相当于放大倍率最高能达到几千倍。

在德布罗意之前，因为理论限制，人类并不奢望能获得放大倍数更大的显微镜，但是，当德布罗意提出实物粒子具有波粒二象性以后，人们很快就意识到，既然电子也具有波动性，如果用电子束来代替光波制作显微镜，其分辨率就能大大提高，因为电子束的波长远远小于可见光，比如受40 kV高压加速的电子，其波长仅为0.006 nm，比可见光小5个数量级。

1926年，德国物理学家布希发现轴对称分布的电磁场具有使电子束偏转、聚焦的作用，与光线通过玻璃透镜的聚焦原理一致，这就意味着用电子束作“光源”，用电磁场作透镜，理论上是可以制造显微镜的。

1932年，德国科学家恩斯特·鲁斯卡（1906—1988）成功制造了世界上第一台电子显微镜。1933年，鲁斯卡研制成功了由多级成像磁

透镜、聚光镜、试样室、高真空镜筒组成的透射电子显微镜（简称透射电镜），成为当今全世界都广泛使用的电子显微镜的先导。鲁斯卡根据量子理论进行了计算，考虑到各种技术上的困难，他预测未来的电子显微镜分辨距离应能达到0.22 nm。

事实上，鲁斯卡还是保守了。现代透射电镜分辨距离可达0.1 nm，已经达到了原子级别，放大150万倍。透射电镜的发展，有力地推动了物理、化学、材料科学、分子生物学、医学等领域的发展，成为人类探索微观物质世界必不可少的技术（图24-1）。

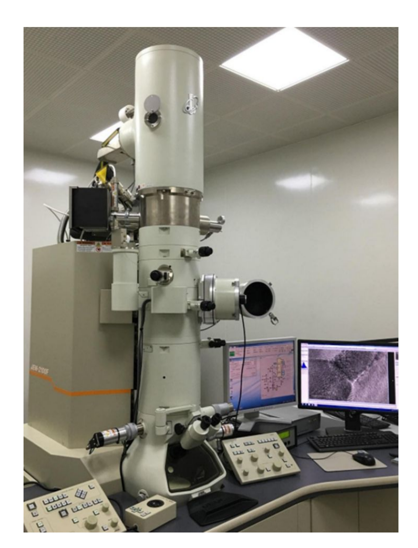


图24-1　透射电子显微镜

(http://www.sim.ac.cn/kybm2016/xxgnclgjzdsys2016/kytp2016/202001/t20200106_5483125.html)

在显微技术的发展历史上，如果说光学显微镜是第一个里程碑，那么透射电镜就是第二个里程碑，而第三个里程碑则是扫描隧道显微镜的发明。扫描隧道显微镜的放大倍数可高达一亿倍，分辨距离达0.01　nm，使人类第一次“看见”了单个原子，是世界重大科技成就之一。扫描隧道显微镜的原理和前两种显微镜完全不同，打个比方来说，如果前两种显微镜是用眼睛看物体表面的话，那么扫描隧道显微镜就是用手在摸物体表面，从而感知表面的凸凹不平。

随着对量子隧道效应研究的深入，人们发现，当两个金属表面非常接近时，施加很小的外加电压（0.002～2 V），电子就会穿过表面空间势垒（两金属间的绝缘层）形成隧道电流。隧道电流有一个奇特的性质：在一定电压下，隧道电流随间距增加而急剧减少，呈指数变化关系。这一变化非常敏锐，距离的变化即使只有一个原子直径，也会引起隧道电流变化1000倍。

人们意识到可以利用这一现象来构建物体表面的微观形貌，其实就相当于是显微镜。但是这需要一个只有几个原子直径大小的探针，而且对探针的控制精度要求极高，与被测物体的距离需要小到1 nm左右，这在技术上是极为困难的。

20世纪70年代，曾有科学家尝试制作这样的显微镜，但是最终失败了。但是到了1981年，终于由IBM苏黎世实验室的格尔德·宾宁和海因里希·罗雷尔制造成功。因为这种显微镜是利用量子隧道效应在物体表面来回扫描，所以被称为扫描隧道显微镜。

扫描隧道显微镜以一个非常尖锐的金属（如钨）探针（针尖顶端只有几个原子大小）为一电极，被测样品为另一电极，在它们之间加上0.002~2 V的电压。当探针针尖在被测样品表面上方做平面扫描时，即使表面仅有原子尺度的起伏，也会导致隧道电流非常显著的变化。这样就可以通过测量电流的变化来反映表面上原子尺度的起伏，从而得到样品表面形貌，如图24-2（a）所示。

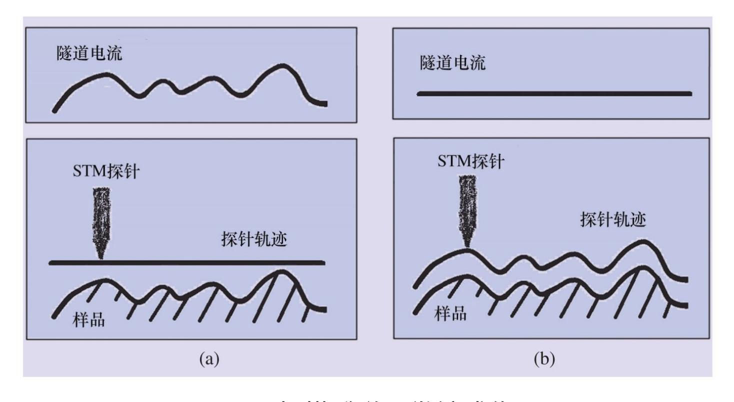



图24-2　扫描隧道显微镜成像原理

（a）探针高度恒定模式；（b）隧道电流恒定模式

还有一种测量方法，通过电子反馈电路控制隧道电流在扫描过程中保持恒定。那么为了维持恒定的隧道电流，针尖将随表面的起伏而上下移动，于是记录针尖上下运动的轨迹即可给出表面形貌，如图24-2（b）所示。

硅片是制造半导体集成电路的主要材料，制造集成电路用的硅片表面必须是高度平整光洁的，所以需要将单晶硅棒切割成一片一片薄薄的硅单晶圆片（简称晶圆）。如果把一块单晶硅切开，最表面的那层原子周围的化学键必然被切断，表面原子就会重新构建化学键，这就是所谓的“表面重构”。沿（111）晶面方向切开的硅的表面出现的重构被称作7×7结构。自1959年发现该结构以来，其原子如何排列一直困扰着人们，也成为热门课题。1983年，宾宁和罗雷尔利用他们发明的扫描隧道显微镜第一次直接观察到这种7×7结构（图24-3），终结了学术上多年的争论，引起了极大的轰动，这是人类第一次亲眼看见原子的真面目。

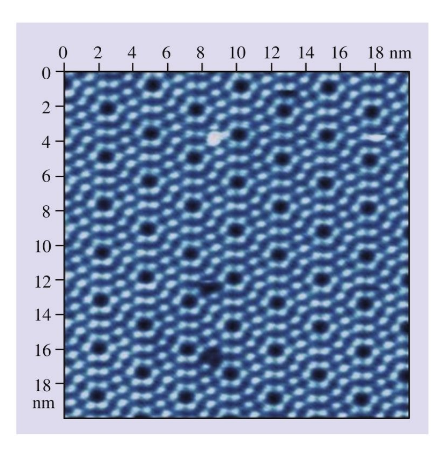



图24-3　Si(111)表面7×7结构图像

(https://www.yiqi.com/product/detail_386152.html)

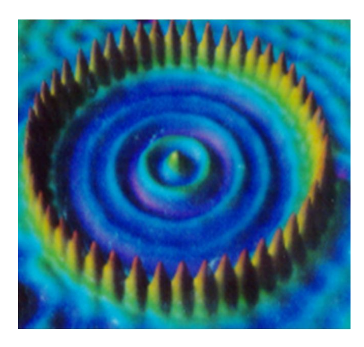


图24-4　48个铁原子形成的量子围栏

扫描隧道显微镜不但可以用来观察材料表面的原子排列，而且还能用来移动原子。可以用它的针尖吸住一个孤立原子，然后把它放到另一个位置。图24-4是IBM公司的科学家精心制作的“量子围栏”。他们在极低的温度下用扫描隧道显微镜的针尖把48个铁原子一个个地排列到一块精制的铜表面上，围成一个围栏，把铜表面的电子圈了起来。图中圈内的圆形波纹就是这些电子的概率波图景，电子出现概率大的地方波峰就高，它的大小及图形和量子力学的预言符合得非常好。

1986年，鲁斯卡、宾尼和罗雷尔3人共同获得了当年的诺贝尔物理学奖。无论是透射电子显微镜，还是扫描隧道显微镜，都是量子物理带给人类最好的礼物。

# 第八篇　量子·前沿技术

# 25 量子计算之算法

20世纪80年代开始，量子技术有了进一步的发展。量子力学从幕后走到了台前，诞生了量子信息技术，如量子计算机、量子密钥传输、量子隐形传态等。这些技术遵从量子力学的运行规律，实现的功能也反映了量子的特性，从而开辟了信息技术的发展新方向。一旦这些技术获得广泛应用，人类社会将再次发生翻天覆地的变化。

2021年，两条重大科技新闻登上了各大媒体的头条，我国研制的光量子计算机"九章二号"（图25-1）处理"高斯玻色取样"问题的速度，比全球最快的超级计算机快上亿亿亿倍；研制的超导量子计算机"祖冲之二号"（图25-2）对"量子随机线路取样"问题的处理速度，比全球最快的超级计算机快1000万倍以上。"九章二号"和"祖冲之二号"的成功，使我国成为唯一在两条技术路线上实现"量子优越性"的国家。

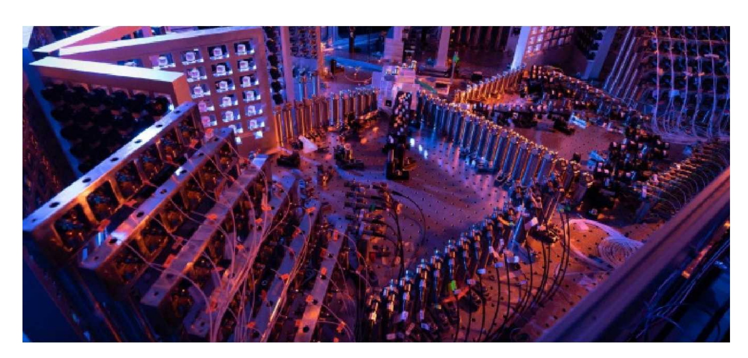


图25-1　"九章二号"

(https://m.thepaper.cn/newsDetail_forward_15074917)

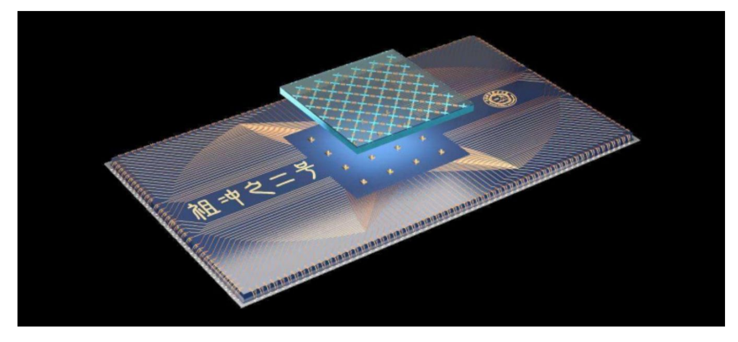



图25-2 “祖冲之二号”

(https://www.sohu.com/a/497277817_ 121119256)

量子计算机为什么有这么大的算力？它和我们的经典计算机到底有什么不同呢？最大的不同，就在于它利用了量子力学的两大特性——叠加与纠缠来实现运算。而与此同时，另外两种量子特性——退相干与测量则成为它的软肋。

人类进入信息时代，以半导体芯片为核心的经典计算机居功至伟。芯片制造简单来说就分为两个大步骤：第1步是在单晶硅上制造几十亿个晶体管，第2步是用导线把这些晶体管按设计好的电路连接起来。晶体管越小、导线的宽度（线宽）越小，芯片集成度越高。28 nm、14 nm、7 nm、5 nm等制程工艺就可以代表导线的线宽，也能代表晶体管的尺寸。早在20世纪60年代，英特尔创始人之一戈登·摩尔就预测同样大小的集成电路上可容纳的晶体管数目每隔18个月便会增加1倍。这样的“神预言”竟然和后来集成电路的发展速度基本吻合，于是就被上升到了定律的高度，称为摩尔定律。

目前，芯片制造工艺已经进入7 nm、5 nm，甚至3 nm阶段，更小尺寸的技术也在研发。但是原子的直径大概在0.2~0.4 nm，也就是说，1 nm相当于3~5个原子排列在一起，如果线宽进一步下降到小于1 nm级别，量子隧道效应将不可避免地影响电子元器件的正常工作。尽管研究人员正在努力通过各种手段进一步延续晶体管的制程尺寸，但是已无法阻止“摩尔定律”必将被打破的历史趋势。因此，研制以量子力学为基础的量子计算机已经是势在必行。

早在1981年，费曼就在一次演讲中指出，用经典计算机来模拟量

子系统的演化存在本质上的困难，其天文数字的计算量是经典计算机无法承受之重。所以他建议用量子体系去模拟量子体系。也就是说，可以构造一个量子体系，其演化的方式跟要模拟的体系在数学上是等价的，然后测量这个量子体系的演化结果，由于结果是概率性的，每测量一次相当于取一次样，多次取样以后我们就知道了这个概率分布。这实际上就是一种量子计算机的模型，事实上，现在有些类型的量子计算机执行的任务叫做某某取样，其思路就来自于此，如上文提到的“高斯玻色取样”和“量子随机线路取样”。

费曼提出量子计算机时，只是希望量子计算机能够帮助科学家解决一些量子力学里的特定问题，并没有指望它能解决经典问题。但是到了1994年，量子计算机出现了一个里程碑式的突破，美国物理学家彼得·肖尔（Shor）发现了一种量子算法——分解质因数算法。肖尔的算法向人们展示，相对于经典计算机，量子计算机可以大幅度提高分解质因数的速度，这立即引起了轰动。人们终于发现，除了量子力学问题，量子计算机还能更快速地解决某些经典的数学问题，极具应用前景，从此掀起了量子计算机的研究热潮。

分解质因数是我们在小学就学过的数学问题，例如，21可以分解成3×7，看起来很简单，但是，如果让你分解291 311，你还能回答出来吗？所以说，这个看似简单的问题其实是一个很难的问题：将两个大质数相乘十分容易，但是想要对其乘积进行因式分解却极其困难。对于那种由两个很大的质数相乘得到的数，经典计算机需要花费大量时间才能把它的质因数找出来。对于经典计算机，如果计算机一秒能做1012次运算，那么分解一个300位的数字需要15万年，分解一个5000位的数字需要50亿年！

因为质因数如此难以分解，所以在保密领域大有用处。现在广泛应用的一种密码协议叫RSA密码协议，就是采用这种手段加密。这个密码协议中，两个质数的乘积是公开的，但这两个质数是保密的，破译者必须将这个乘积分解为两个质因数才能破译密码。例如，现在RSA密码协议中需要破译的整数用二进制表示有2048位，为了破解这个密码，量子算法大约需要$1.6\times10^{8}$步，而经典算法则需要大约$6.75\times10^{51}$步。假设量子计算机和经典计算机每秒都能算$10^{9}$步，那么

量子计算机不到1 s就能破解密码，而经典计算机则大约需要$2\times10^{35}$年，这简直是降维打击！

但是，读者不要高兴得太早，我们上述假设的基础是量子计算机每秒能算$10^9$步，而目前量子计算机的硬件研发还处于初级阶段，还没法实现这样的计算能力。目前，公开报道的最佳性能，是我国科学家于2017年用量子计算机成功分解291 311这个数字（291 311 ＝523 ×557），291 311换算成二进制，是一个19位数，对于经典计算机，分解这个数字也是轻而易举的事情。

那么，量子计算机为什么这么难造？它到底是怎么制造的呢？

现代计算机都是采用二进制的“比特”（也叫“位”，用“0”或“1”表示）作为信息单位，工作时将所有数据排列为一个比特序列，进而实现各种运算。对于经典计算机而言，通过控制晶体管电压的高低电平，就可以决定一个比特到底是“1”还是“0”，高电平代表“1”，低电平代表“0”。为了避免各种干扰的影响，高低电平并不是一个固定的值，而是一个变化范围，只要在这个范围之内，就可以区分开这两种状态，所以比较容易控制（图25-3）。

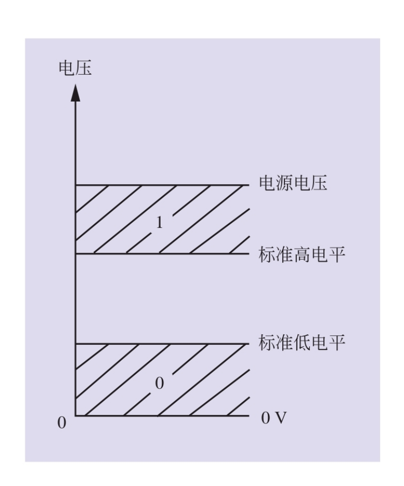



图25-3　经典比特高低电平的确定

而量子计算机使用的是量子比特，能秒杀传统计算机得益于两个独特的量子效应——量子叠加和量子纠缠。量子比特最大的特点，是

它可以处于“0”和“1”的叠加态，即一个量子比特可以同时具有“0”和“1”两种状态。显然，如果有$n$个量子比特，它们纠缠在一起，就能创造出一种超级叠加，这时它们的组合就有$2^n$个状态。对这样的状态进行一次操作，就相当于对$n$个经典比特进行了$2^n$次操作。也就是说，$n$个量子比特的计算能力是$n$个经典比特的$2^n$倍。由此可见，量子计算机的计算能力可随着量子比特位数的增加呈指数增长，这是一个惊人的增长速度，这一特性让量子计算机拥有超强的计算能力。

如果读者听过一个小故事，就会对指数增长有更深刻的认识。说有一个农民与国王下国际象棋，国王说你如果赢了奖赏随便要，农民说：“我只要在棋盘的第一格放一粒米，第二格放二粒，第三格放四粒，第四格放八粒……每次都翻一倍，放满这棋盘的64个格子就行了。”国王一听，哈哈大笑，心想农民真是没见过世面，米如果一粒一粒数，装几麻袋顶天了，就答应下来。可是，当国王输了以后，让人把米扛来，才发现这一次可输大了，整个国家的米都不够用！所有64个方格上的米粒总数为：$1+2+4+8+\dots+2^{63}$，算下来大概是$1.8\times10^{19}$粒，约18亿t，这国王如何赏得起？由此可见指数增长的惊人威力。

所以说，如果量子计算机的量子比特不多，其威力并不明显，例如，一台由10个量子比特组成的量子计算机，其运算能力相当于1024位的传统计算机。但是，如果量子计算机拥有50个量子比特，其性能就能超过世界上绝大多数超级计算机，如果拥有300个量子比特，就能将世界上最先进的超级计算机需要数万年来处理的运算缩短至几秒钟。

但是，成也萧何，败也萧何。量子叠加可以使量子计算机执行惊人的运算速度，但麻烦的是，由于运算过程处于叠加态，所以运算结束后也是叠加态，要想得到运算结果，必须进行测量，可是测量结果只有一个，也就是说，本来有$2^n$个数据，你一测量，就剩一个了，其他全没了。为了得到其他数据，你不得不重复所有的计算。很显然，为了得到所有的结果，重复计算的次数不会比所需结果的数目少，这样看来，量子计算并不会比经典计算更节省时间。也就是说，简单应

用态叠加原理并不会使量子计算机获得计算优越性。

那么，如何才能利用量子计算的巨大潜力呢？很简单，如果对于某些计算问题，不需要获得所有的计算结果即可解决问题，那不就行了吗？行是行，但是，这就需要进行非常巧妙的算法设计。目前，只有少数问题人们获得了高效的量子算法，而对于绝大多数问题，如简单的加减乘除，还没有相关算法。

如肖尔提出的分解质因数算法，肖尔发现，在计算结果中存在某种周期性规律，这样，我们就不需要获得所有结果，只要找到这个周期性规律就能间接实现质因数分解。在肖尔算法中，通过对输入的量子态进行傅立叶变换操作是算法的核心，这是一种非测量性的变换操作，能将所寻找的周期值转移到单个测量结果中，这是减少测量操作的关键。由于傅立叶变换本身的操作比测量出全部计算结果的操作能节省大量时间，所以这种方法比经典计算能实现指数级加速。

1996年，美国科学家格罗弗（Grover）发现了另一种很有用的量子算法——量子搜索算法。它可以在一个海量的无序的数据库中寻找某些符合特定条件的元素。这个算法虽然达不到指数加速，但是可以把搜索问题从经典的$N$步缩小到$\sqrt{N}$步，从而显示出量子搜索的优越性。这个算法的特点是，利用不同状态间的相干性，设计出合理的量子算法，使得通往正确状态的概率能够迅速叠加增长，经过若干次重复运行后正确状态的概率就能趋近于1。此时进行测量，结果即为正确结果。例如，电话本以号码排序，共有个100万个号码，要从中找出某人的电话号码。经典方法是一个个找，平均要找50万次，才能以50%的概率找到所要的电话号码。而量子算法每查询一次就可以同时检查所有的100万个号码，由于量子比特处于纠缠态，量子干涉效应会使前次的结果影响到下一次的量子操作，这种干涉生成的操作运算重复1000次后（即$\sqrt{1\ 000\ 000}$），获得正确答案的概率为50%，如果再多重复操作几次，则可以以接近于1的概率找到所需的电话号码。

2008年，量子算法又取得突破。麻省理工的3位科学家开发了一种求解线性方程组的量子算法，被称为HHL算法。这种算法并不能全方位代替经典的求解线性方程组的算法，只有当求出的线性方程组的

解不需要读出（这就省去了测量的麻烦），而只是作为其他算法的输入值的时候，HHL算法才有可能提供计算加速。这一算法在量子机器学习中有很好的应用前景。

现在，人们已经开发出几十种量子算法。随着量子算法的研发，量子计算机的硬件研发也迅速跟进，关于量子逻辑门、量子电路等许多设计方案不断涌现，使得量子计算的理论和实验研究蓬勃发展。

# 26 量子计算之硬件

有了量子算法以后，人们需要做的，就是如何造出一台量子计算机。一个量子计算机的工作原理分成四步：（1）构建可以表示量子比特的物理体系；（2）把所有量子比特初始化为一个给定的量子态；（3）对这些量子比特进行一系列逻辑操作，控制和操作量子态的演化和传递，最终到达某个量子态；（4）对最后的量子态进行测量，读出结果。其工作原理如图26-1所示。图中的逻辑操作是信息处理的核心，首先选择适合于待求解问题的量子算法，然后将该算法按照量子编程的原则转换为控制量子芯片中量子比特的指令程序，从而实现逻辑操作的功能。用的逻辑操作的步数越少，算法就越快。

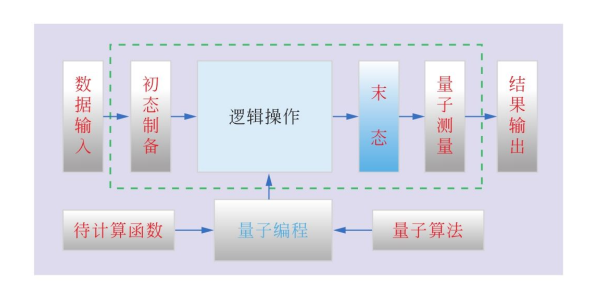



图26-1　量子计算机的工作原理

显然，要想造出一台量子计算机，第一步构建可以表示量子比特的物理体系是最基础的。我们前面谈到，经典比特的“1”和“0”通过控制晶体管电压的高低电平来实现，那么，对于量子比特，需要用什么物理体系来实现“0”和“1”的叠加态呢？事实上，可供选择的体系非常多，如我们最熟悉的体系——光子的偏振态就可以。把光子的垂直偏振态作为“1”，水平偏振态作为“0”，那么每一个光子就可以作为一个量子比特。除偏振叠加态之外，还可以采用光子的路径叠加态以及其他一些自由度的叠加态来构建量子比特，而且实现多个光子比特纠缠的技术也比较成熟。我国的潘建伟院士团队在光量子计算机的研制方面一直处于世界领先地位，在该领域取得了多项突破：2007年，团队

用四个光子比特成功实现了肖尔算法，演示了数字15的质因数分解；2013年，团队基于光量子计算平台成功实现线性方程组量子求解算法，求解了一个2×2的线性方程组；2017年，团队成功研制出五光子玻色采样计算原型机（图26-2），在采样率上首次超越早期经典计算机；2021年，“九章二号”（图25-1）横空出世，实现了113光子144模式的玻色取样，多光子量子干涉线路达到了144维度，探测到的光子数达到了113个，均刷新了世界纪录。

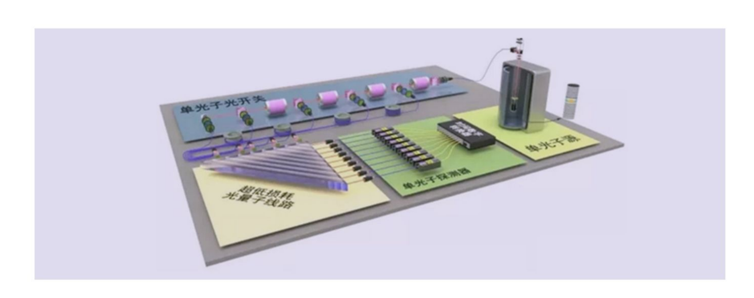



图26-2　中科院研制的光量子计算原型机结构图

理论上，任何处于叠加态的粒子或处于叠加态的量子状态都可以作为量子比特，所以除了光子以外，量子计算机常用的物理体系还有离子阱（被囚禁的离子）、超导约瑟夫森结、超冷原子（接近绝对零度的原子）、金刚石色心（钻石中的一种晶格缺陷）、半导体量子点（量子点指的是尺寸在纳米级的材料）等。这些物理体系里，又有多种方式构建量子比特，如超导体系就可以分为电荷量子比特、磁通量子比特、相位量子比特等类型（这些类型的叠加态都比较复杂）。再如半导体量子点体系可分为电荷量子比特（电子位置在左和在右的叠加态）和自旋量子比特（电子自旋向上和向下的叠加态）等类型。

相对于其他物理系统，超导量子计算机在各种技术路线中被寄予厚望，这是因为基于超导量子电路的量子计算有以下优势：（1）超导量子电路是一种电路，有很高的设计自由度；（2）超导量子比特的操控使用的是工业上广泛应用的微波电子学设备，易于实现复杂的调控；（3）超导量子芯片的制备工艺是基于成熟的半导体芯片微纳加工技术，相对容易扩展到由大量比特构成的复杂芯片。

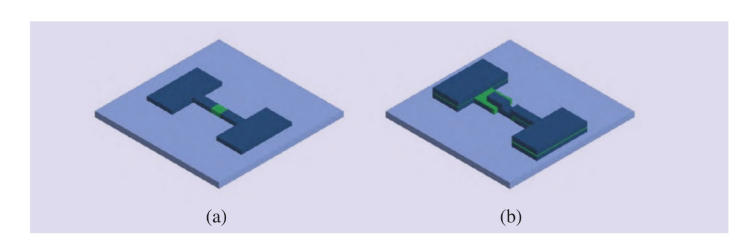


图26-3　两种常见类型的约瑟夫森结示意图（图中深蓝色为超导体，绿色为绝缘层）

超导体系的核心物理器件是约瑟夫森结，这是一种“超导体-绝缘体-超导体”的三层结构（如$Al$-$Al_2O_3$-$Al$），其中绝缘层厚度只有几纳米（图26-3），在超低温下可表现出宏观量子效应。在两块超导体之间夹一个绝缘层，按经典理论，电子是不能通过绝缘层的。但在1962年，英国物理学家约瑟夫森根据隧道效应从理论上做出预言，只要绝缘层足够薄，超导体内的电子就可以通过隧道效应穿过绝缘层而形成电流。1963年，实验证明了约瑟夫森预言的正确性。利用超导约瑟夫森结来观测宏观量子效应最早在1985年提出，随后研究人员在超导约瑟夫森结器件中陆续观测并实现了能级量子化、量子隧穿、量子态叠加、量子相干振荡等现象，为超导体系打下基础。

美国的谷歌公司在超导量子芯片方面多年来处于世界领先地位，图26-4是其2019年推出的53个量子比特的超导量子芯片，该量子计算机被命名为“悬铃木”。我国的超导量子计算机发展速度也很快，2021年，由潘建伟团队研制的“祖冲之二号”（图25-2）超导量子计算机包含66个量子比特，超越了“悬铃木”。

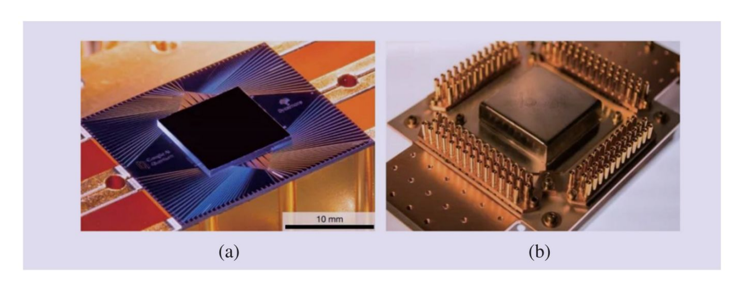



图26-4　谷歌推出的53个量子比特的超导量子芯片

（a）量子比特处理器；（b）封装好的芯片

除了超导量子计算机以外，基于半导体量子点的量子计算机也可以结合现代半导体微电子制造工艺来制造，也是最有希望的候选者之一。这一技术路线最早在1998年提出。

我们知道，经典计算机芯片依赖于晶体管，随着摩尔定律的发展，晶体管尺寸越来越小，那么，当晶体管小到极限以至于只能容纳一个电子时，那会是什么情况呢？这就是半导体量子点，有时也称为单电子晶体管（图26-5）。2018年，我国郭光灿院士团队以单个电子的量子点作为量子比特，创新性地制备了半导体六量子点芯片，在国际上首次实现了半导体量子点体系中的三量子比特逻辑门操控，为未来研制集成化半导体量子芯片打下基础。

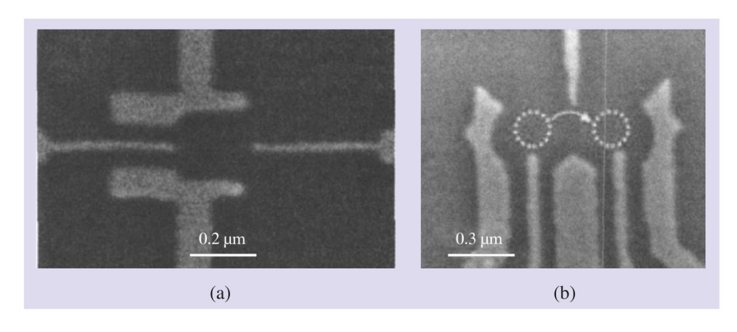
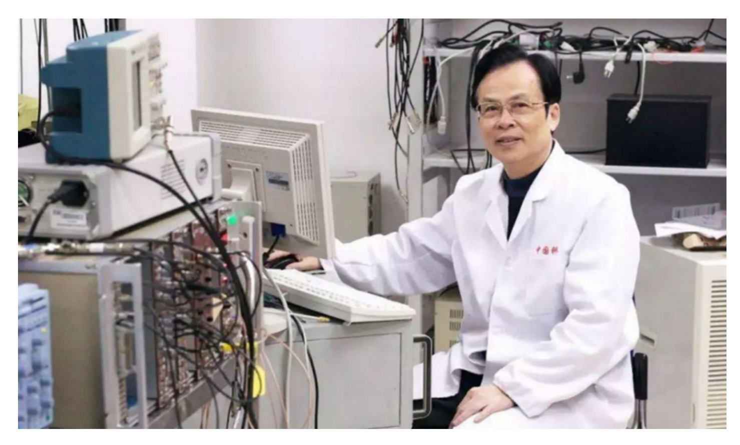



图26-5　半导体量子点的电极结构图

（a）典型的单量子点结构；（b）典型的双量子点结构



郭光灿（中国科学院院士）

(http://www.kicchina.org/index.php?m=content&c=index&a=show&catid=17&id=105)

看到现在，相信读者已经对量子计算机有了一个初步的了解，所以大家再看到关于量子计算机的新闻时，就要重点关注三个问题：第一，它执行的是什么量子算法；第二，它用的是什么物理体系；第三，它有多少个量子比特。如前所述，量子计算机的算力是随着比特数的增加呈指数上升的，所以量子比特的数目非常重要，它决定了量子计算机的性能上限。但是，在工程上提升比特数目是一件很困难的事情，量子比特越多越难造。

量子计算机发展到现在，还没有进入实用阶段，因为从理论到工程面临着众多棘手难题，其中最主要的一点就是经典计算机根本不存在的问题——退相干。

根据退相干理论（见第14章），当量子体系与外界环境相互作用后，就会发生退相干过程，使量子体系逐渐退化为经典体系，失去量子特性。量子计算机是宏观尺度的量子器件，环境噪声和逻辑操作不可避免地会导致量子比特相干性的消失，使叠加态逐渐退化为确定态，这样，量子计算机就退化成了经典计算机，失去了其使用价值。

所以，衡量某种技术路线的量子计算机的发展前景，有一个很重要的指标就是退相干时间。表26-1给出了各种物理体系的基本指标比较。

表26-1　量子计算机各种物理体系的基本指标比较

| | 光子 | 超导电路 | 半导体量子点 | 离子阱 | 金刚石色心 | 超冷原子 |
|---|---|---|---|---|---|---|
| 退相干时间 | 长 | ~10 μs | ~10 μs | >1000 s | ~10 ms | ~1 s |
| 可扩展性 | 较好 | 较好 | 较好 | 较差 | 较差 | 较差 |
| 运行环境 | 常温 | 极低温 | 极低温 | 极低温 | 常温 | 极低温 |

注：退相干时间都是现阶段水平，将来会不断有新的突破。

退相干时间指的是量子相干态演化到经典状态的时间。量子计算

必须在叠加态上进行，否则量子运算就没法持续下去，因此，退相干时间越长越好。为了尽量减小环境对相干性的影响，量子计算机对环境要求相当苛刻，大部分体系都需要在极低温（接近绝对零度）和超高真空的环境中运行。即便如此，环境还是有干扰，量子态还是非常“脆弱”，因此，人们不得不采用量子编码来纠错。

“量子编码”包括量子纠错码（出错后纠正）、量子避错码（应用量子相干保持态避免出错）和量子防错码（多次测量防止出错）等。量子纠错码是发明分解质因数算法的肖尔在1995年提出的，量子避错码是郭光灿院士团队在1997年提出的。

量子纠错码用于纠正环境退相干造成的错误，是目前研究的最多的一类编码。它是从经典纠错码类比得来，其优点为适用范围广，缺点是效率不高。

我们先了解一下经典纠错码的技术实现。如前所述，经典计算机通过控制晶体管电压的高低来决定一个比特是“1”还是“0”。虽然“1”或“0”都对应一个较大的电压范围，但在噪声的扰动下，一个处于0态的比特还是有很小的可能变成1，导致错误。为了尽可能避免和减少错误，经典纠错方案是把3个比特当作1个比特用：

$$000 \rightarrow 0 \qquad 111 \rightarrow 1$$

通常把左侧3个比特叫做物理比特，把右侧的1个比特叫做逻辑比特，其中逻辑比特是信息处理的单元。3个物理比特处于000代表逻辑比特处于0，3个物理比特处于111代表逻辑比特处于1。假设由于噪声，处于000态的物理比特变成了010，由于2个比特同时出错的概率很小，计算机就判定是中间的物理比特出错，实施操作将其纠正为000，这就降低了错误率。

量子纠错码与之类似，用若干物理量子比特来编码1个逻辑比特，用以纠正退相干引起的错误。不同的是，量子编码需要用更多的物理比特来纠错。业已证明，至少需要5个物理比特编码，才能实现1个逻辑比特的纠错。可以说，这既是一个好消息，又是一个坏消息。好消息是，量子计算机可以制造；坏消息是，它极大地增加了制造的难度。量子计算机至少需要50个逻辑比特才有可能超越经典计算机，

这就至少需要250个物理比特，加起来达到了300个量子比特的规模，这就对物理体系的可扩展性提出了极高的要求。

可扩展性指的是系统量子比特数目的扩展。和经典计算机的简单并列就可以增加比特不同，量子计算机需要量子比特都纠缠在一起并准确操控，因此每增加一个比特都极为不易。而且，量子比特的数目越多，退相干就越容易发生。因此，集成300个量子比特面临着非常大的技术挑战，目前的最高纪录也与之相去甚远，我们距离具有实用价值的量子计算机还有很长很长的路要走。

另外，量子计算机目前还有一大缺点是没有内存。经典计算机的调试依赖于内存和中间计算机状态的读取，这在量子计算机中是不可能的。量子状态不可以像经典计算机那样简单复制以供以后检查，对量子状态的任何测量都会将其坍缩为一组经典比特，从而使计算停止。因此，新的调试方法对于大规模量子计算机的开发至关重要。2021年，郭光灿院士团队打造出“量子U盘”，可以将光信息存储在特殊晶体中1 h，大幅刷新德国团队创造的1 min的世界纪录，具备了实用化的前景，这也为量子计算机构建内存带来希望。

展望未来，科学家们并不满足于只能执行特定算法的专用量子计算机，他们的终极目标，是制造可编程的通用量子计算机，可以用来解决所有可计算问题，可在各个领域获得广泛应用。通用量子计算机的实现必须满足两个基本条件：一是量子比特数要达到几百万量级，二是应采用纠错容错技术。鉴于目前量子计算机的研制还处在初级阶段，因此通用量子计算机还只是理论上的蓝图，距离我们还很遥远。不少物理学家认为，通用量子计算机从蓝图变为现实可能需要50年甚至更长的时间。征途漫漫，唯有奋斗。

# 27　量子密码

2016年，一条科技新闻引发了全球的关注，人类历史上第一颗量子科学实验卫星“墨子号”成功进入太空。该卫星由潘建伟团队牵头研制，运行在高度约500　km的近地轨道，是世界第一颗探索太空与地面量子通信可行性的卫星。升空之后，“墨子号”配合位于河北、青海、新疆、云南等地的多个地面站，成功进行了星地量子密钥分发、星地量子纠缠分发以及星地量子隐形传态等实验。图27-1是“墨子号”与地面观测站建立星地量子信息通道的场景，犹如科幻大片般梦幻，美不胜收。截至目前，“墨子号”依然是世界上唯一在轨的具备量子通信终端能力的卫星。



图27-1　采用延时摄影拍摄的“墨子号”建立星地量子链路过程（红色为地面信标光，绿色为星上信标光）

“墨子号”最主要的功能是实现了高速的星地量子密钥分发。近年来，在地面上通过光纤运行的量子密钥分发技术日渐成熟，已经实现了产业化，但传输距离仍然是其短板，而“墨子号”就是为了补齐这一短板，实现超远距离传输，为构建天地一体化量子通信网络探路而研发的。

我想读者朋友们现在一定很好奇，到底什么是量子密钥分发呢？在了解量子密钥之前，我们需要先简单了解一下通信与密码学。

我们日常传输的信息是由符号、文字、图像、语音等构成的，但在现代计算机和通信系统中，这些信息都被表示成由0和1构成的比特

串，例如，中文字符“汉”用Unicode编码转换成二进制后得到的比特串是11 100 110 10 110 001 10 001 001，所以通信过程只要传递这个比特串即可。

密码学的基本思想是对数据进行伪装以隐蔽信息，所谓伪装就是对数据进行一组可逆的数学变换，伪装前的原始数据称为明文，伪装后的数据称为密文，伪装的过程称为加密。把明文变换成密文，需要两个元素：加密算法和密钥。加密算法就是变换的规则，密钥就是变换的参数。

下面举个例子来说明现代通信的加密过程。

假设要传递的明文是：00 101 001。

首先设计一个加密算法：设置密钥长度与明文一样，密文由明文每个数字与密钥对应数字相加得到，规定0＋0＝0，0＋1＝1，1＋0＝1，1＋1＝0；

然后随机生成一个密钥：假设为10 101 100；

加密过程如下：

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| 明文 | | 0 | 0 | 1 | 0 | 1 | 0 | 0 | 1 |
| 密钥 | + | 1 | 0 | 1 | 0 | 1 | 1 | 0 | 0 |
| 密文 | | 1 | 0 | 0 | 0 | 0 | 1 | 0 | 1 |

这样就得到了用加密算法加密后的密文：10 000 101。

接收方接收到密文后，通过密钥反向运算即可解锁密文，获得明文信息。事实上，上述加密算法的一个突出优点就是其加密运算与解密运算是一样的，密文与密钥直接相加就可以得到明文。这样，加密和解密可以共用一个软件或硬件模块，使工程制造量减少一半。

在通信过程中，默认为密文和算法都是可以被敌方破解的（因为敌方即使破解了也不会告诉你，你必须假设敌方已经破解），唯一需要绝对保密的就是密钥。所以，发送方如何将密钥安全送达接收方就成为保密通信成败的关键。你可能要问了，如果敌方从你的多次通信中反推出密钥怎么办？现代密码学家早已想到了这一问题，早在20世

纪40年代，信息学鼻祖香农就证明，如果密钥随机生成且长度与明文一样，而且密钥一次一换，绝不重复使用，则这种密文是绝对无法破译的，这就是著名的“一次一密”。

但是，“一次一密”虽然安全，密钥传输却成了大问题。因为“一次一密”要消耗大量密钥，需要甲乙双方不断地更新密码本，这时候，甲方印一本密码本送给乙方的方式肯定不实用了，只能通过光纤、无线电波等现代通信网络传输，而这些信道都有被敌方窃听的可能，而且即使被窃听了你也很难发现。所以，目前的经典通信使用“一次一密”并不广泛。

目前，经典通信广泛使用的方法主要是“公钥加密法”，最常用的是RSA密码协议。这种方法之所以安全，是因为采用了大数分解质因数这种经典计算机无法计算的数学问题。然而，量子计算机的分解质因数算法可以轻易破解这一难题，一旦量子计算机的研究达到实用化，RSA公钥体系将无密可保，那时候该怎么办呢？

唯一的办法，就是放弃RSA加密，找别的加密手段。这时候，“一次一密”重新进入了人们的视线，这是绝对无法破译的加密手段，如果能保证密钥传输不被窃听就是绝对安全的。于是，量子密钥就应运而生了。人们发现，量子通信具有一个天然的优势，因为量子测量的随机性，它可以产生绝对随机的字符串，这些字符串是绝佳的密钥。而且因为量子态的不可克隆性和不确定性，任何企图窃取传送中的量子密钥的行为都会被合法用户发现，也就是说，它是没法被窃听的！

首先想到将量子力学用于保密的是美国哥伦比亚大学的一个研究生威斯纳（Wiesner），他在1970年提出一个异想天开的概念——量子防伪钞票。他想象出一种可以在上面保存20个光子的钞票，每个光子都由银行随机用“十字”和“交叉”两种方向的偏振片测量（图15-2），每张钞票的测量结果都保存在银行的数据库里。例如，图27-2就是这样一张钞票，20个光子的偏振状态如图所示，银行数据库了记录了这张序号为A123456的钞票的光子偏振信息。

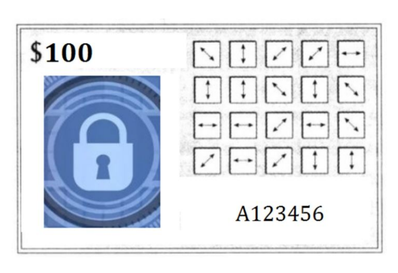


图27-2　威斯纳的量子防伪钞票

如果有人想造假钞，就要面对一个很大的问题——他可以印刷序号A123456，但无法得知这20个光子的偏振状态，他需要测量。但是，他不知道每一个光子是用“十字”还是“交叉”偏振片测量的，所以，一旦他拿错了偏振片，测量结果将发生错误，同时光子的偏振态发生改变（例如，一个45°偏振光子通过十字偏振片，变成水平偏振态或垂直偏振态的概率各一半，如图27-3所示），于是光子偏振信息就不可能正确复制了。这样，银行很容易就能检验出他做的是一张假钞。

这真是一个脑洞大开的想法，但是，即使在现在，想要把20个光子保存到一张小小的纸币上也是天方夜谭。所以，当威斯纳把他的想法写成论文投稿后，杂志社的编辑认为这个年轻人简直就是在胡言乱语，直接退稿。他又投稿到另外三家杂志社，无一例外地全部退稿。同时，他的导师也并不看好他的这一创意，对他的想法不感兴趣。于是，他只好把论文束之高阁。

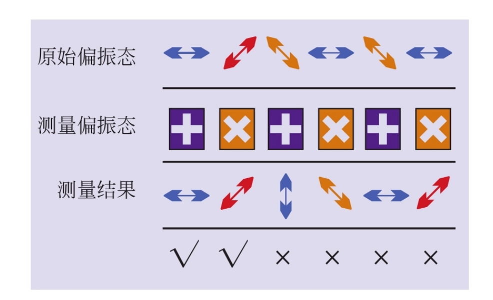


图27-3　测量基对测量结果的影响（后面四个光子选错了偏振片，光子的偏振态发生改变，测量结果全部错误）

1983年，威斯纳终于找到了一个机会，在一个关于密码学的国际会议上发表了这篇论文，这距离他提出这个创意已经过去13年了。不过，赶得早不如赶得巧，恰好参加这次会议美国的密码专家贝内特（Bennett）和加拿大的密码专家布拉萨德（Brassard）对威斯纳的量子防伪钞票很感兴趣，他们很重视威斯纳的创意，并从中深受启发。他们认识到，威斯纳的单光子虽然不好保存，但可用于传输信息，由此可以建立量子密码。经过一年的研究，两人在1984年提出了用单光子偏振态编码的第一个量子密码术方案，现在称之为BB84协议，这便是量子密码的起源。

BB84协议解决的是通信双方的密钥传递问题。经典的密钥传递是甲方预先设定好密钥，然后传递给乙方。而量子密钥并不是预先就有的，它是在甲乙双方建立通信渠道之后，通过双方的一系列量子操作，直接在双方手里产生的，而且不用看对方的数据，就能确定对方的密钥和自己的密钥完全相同。也就是说，量子密钥是一个双向产生的过程，这就好像有一个不存在的第三方把密钥分发给甲乙双方，所以称为“量子密钥分发”。量子密钥分发能使通信的双方产生并分享一个随机的、安全的密钥，这是经典通信不可能完成的任务。

看到这里，有的读者可能会想到，量子纠缠不就能达到这个效果吗？是的，没错，如果用纠缠源产生一对对的纠缠光子，分别发送给甲方和乙方，当他们使用相同的测量基来测量他们各自获得的光子的偏振态时，他们的测量结果是一致的，这样双方就都获得了密钥，这也是后来提出的Ekert91协议和BBM92协议的基本原理。但是，目前纠缠分发的速度还不够快，很难达到实用化的水平，所以，在众多量子密钥分发协议中，研究最深入、实用化程度最高的还是BB84协议，它已经成为目前国际上使用最多的量子密钥方案，并成为量子通信发展的重要基础。

BB84协议是利用单光子来进行量子密钥分发的，下面我们来简单了解一下它的基本实施过程，读者可以将其与威斯纳的量子防伪钞票进行对比，体会二者的区别与联系。

如图27-4所示，在BB84协议中，甲方用单光子源产生一系列光子，并将这些光子通过沿正向或斜向放置的偏振棱镜随机制备成偏振

方向为0°、45°、90°或135°的单光子序列，然后通过量子信道（如光纤或自由空间等）将这些光子传送到乙方，乙方随机选择“十字”或“交叉”检偏棱镜进行测量，将测量结果记录下来。为了方便说明，我们举个例子，假如说甲方发送了12个光子，这些光子的偏振态如图27-5所示，代表着110 001 001 010这样一个字符串，乙方随机测量后，得到的结果是011 001 101 010，显然二者不一样。那该怎么办呢？注意，重点来了。这时候，乙方用经典信道公布所用的测量基（无需保密），甲方告诉他哪些测量基选对了（无需保密），即图中打对号的测量基。这样，双方可以确保打对号的光子测量结果是一致的，于是就保留对的，舍弃错的，这样就得到了密钥1 001 110，然后甲方根据这个密钥用经典通信来传送密文，乙方用这个密钥来解密。

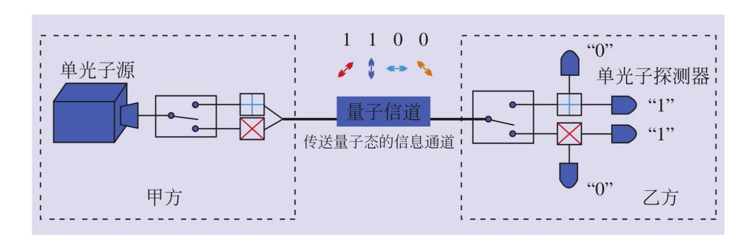



图27-4 BB84协议量子密钥分发过程

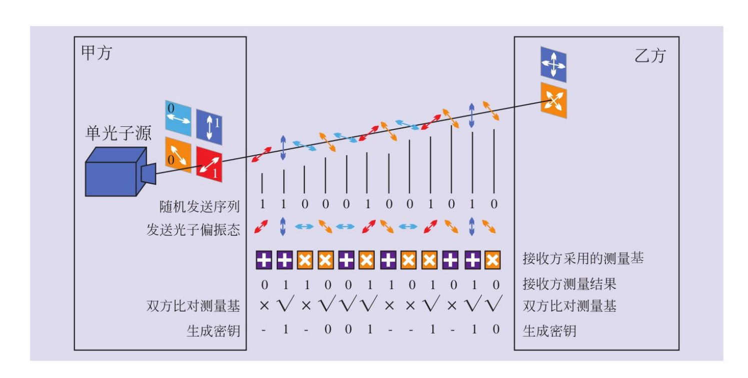



图27-5 BB84协议量子密钥分发过程

(https://baike.baidu.com/tashuo/browse/content?id＝9b9d9c1a6b295335d8269e23&fr＝vipping)

有读者可能要问了，乙方直接公布所用的测量基，甲方告诉他哪

些选对了，都无需保密，不怕敌方知道吗？这就是量子保密通信的妙处了，即使敌方知道了也没用，因为每个测量基都对应着0和1两种测量结果，是0是1只有甲乙双方知道，别人是没法得知的。如果敌方想窃听，只能破坏量子信道，这会导致甲乙双方最终形成的密钥不一致，甲乙随机选择一段密钥进行比对，只要发现误码率异常得高，便知有窃听者存在。

1996年，科学家们给出了BB84协议的严格安全性证明，证明密钥分发过程中只要有人窃听，一定会对体系产生扰动从而被通信双方得知。但是有一个前提条件，就是必须保证每次只发射一个光子才能绝对安全，一次多于一个光子就可能被窃听。而现有的单光子源技术还不成熟，很难投入实际应用，不得不使用一些替代光源，例如，激光经过衰减后得到的弱激光脉冲，而这种激光脉冲每次发射的光子数是不确定的，可能是一个，也可能是多个，这就使窃听者有了可乘之机。好在在2003—2005年，美国西北大学的黄元瑛和我国清华大学的王向斌等提出了诱骗态协议，克服了不完美单光子源带来的量子通信安全漏洞，使得量子密钥分发获得了真正的应用价值。很快，量子密钥分发在光纤中的安全传输距离就突破了100 km，随后，包括我国在内的世界各国开始纷纷布局和推进量子保密通信的实用化。

2017年9月，世界首条量子保密通信网络——“京沪干线”正式开通。京沪干线利用的核心技术就是诱骗态BB84理论方法，该网络在北京、济南、合肥、上海等地的内部量子网络的基础上，通过几十个中继节点把它们连接起来，从而在2000 km的范围内实现量子保密通信。

光纤网络中信号损耗较大，所以需要大量的中继节点才能实现远距离量子通信，而如果借助自由空间来传输信号，损耗就小得多，这样就能实现更远距离的量子通信，这就是“墨子号”量子卫星的优势所在。“墨子号”轨道高度为500 km左右，只有在10 km的大气层内有信号损耗，出了大气层接近真空，信号基本不会受到影响，因此大大拓展了传输距离。然而，卫星与地面之间建立信号通道的困难也是显而易见的，卫星相对于地面以每秒几千米的速度掠过，单光子信号又非常微弱，所以双方对准探测器非常困难，打个比方来说，其精度相当

于在50 km外把一枚硬币扔进一列全速行驶的高铁上的一个矿泉水瓶里，而且为了保证卫星与地面站的通信，卫星过站期间必须一直保持这种精确的通信连接状态，其难度可想而知。

令人难以置信的是，“墨子号”居然做到了。“墨子号”在经过地面站的时间段内，卫星上量子诱骗态光源平均每秒发送四千万个信号光子，一次过轨对接实验可生成300 KB的密钥，平均成码率可达1.1 KB/s，已经初步具备了实用功能。但由于“墨子号”是低轨卫星，相对地面飞行速度较快，每次过站时间小于10 min，并且采取了夜间工作模式来避免阳光的干扰，因此还无法满足全天候的通信需求。

2017年，“京沪干线”与“墨子号”成功对接，这标志着我国已构建出全球首个天地一体化的广域量子通信网络雏形。科学家们未来的目标，是发射多颗由高轨卫星和低轨卫星共同组成的“量子星座”，与地面光纤网络一起，打造真正的“量子互联网”。

# 28 毁灭与重生

1999年，自然科学领域的顶级期刊《自然》（*Nature*）精选了一百多年来该杂志所发表的21篇物理学论文，组成特刊“百年物理学21篇经典论文”，以此纪念百年来物理学所取得的伟大成就。这些论文里，包括伦琴发现X射线、爱因斯坦介绍相对论发展、沃森和克里克发现DNA双螺旋结构等重要论文，而令人瞩目的是，其中竟然有一篇仅仅发表2年的论文——《量子隐形传态实验》。这篇论文是奥地利的塞林格团队（潘建伟是该论文的第二作者）发表的，他们成功地在世界上首次实现了量子隐形传态。1997年，该论文一经发表就引起了轰动，成为量子信息领域的经典之作。

你一定很好奇，什么是量子隐形传态？它到底有什么神奇的魔力，能让世界为之瞩目？

隐形传送，可以说是人类长久以来的梦想，一个人在某处神秘消失，而后又在另一处神秘出现，这是不少科幻小说中出现的场景。这种场景非常令人神往，但人们也都知道，这不过是科学幻想罢了。而量子隐形传态的出现，则让人们似乎看到了一丝希望。科学家们提出的“量子隐形传态”方案，可以使粒子的量子态在某处消失，随后在另一处重现，真的有点像科幻中的隐形传送。

但是，量子隐形传态和科幻中的隐形传送还不太一样。我们想象中的隐形传送是把一个粒子从甲地传送到乙地，而量子隐形传态则是将甲地的某一粒子的未知量子态在乙地的另一粒子上还原出来。也就是说，甲地的粒子并没有移动，它还待在原地，不过，它的“灵魂”被转移到了乙地的另一个粒子身上，那个粒子变得和它一模一样，就像把它传送过去一样。所以，这里头有个词很关键，叫“传态”而不是“传送”，“传送”是直接传送粒子本身，而“传态”只是传送量子状态。

那么，这是不是相当于在乙地复制出了甲地的粒子呢？还不能叫复制，因为复制过程原件并不会损坏，而在量子隐形传态过程中，必须把“原件”摧毁才能获得“复制件”，因为从理论上来讲，如果不损

坏“原件”，量子态是不可复制的，这是由“量子态不可克隆原理”决定的。1982年，物理学家从态叠加原理得出推论，对任意一个未知的量子态进行精确的完全相同的复制是不可实现的，这就是“量子态不可克隆原理”。其实这并不难理解，“克隆”是在不损坏原有量子态的前提下再造一个相同的量子态，而任何一个量子态都是处于叠加态的，想克隆它就得对它进行测量，一测量就会变成确定态，它就被破坏了，你如何能克隆它呢？

不可克隆，那就想别的办法。1993年，贝内特（Bennett）等6位科学家联合发表了一篇题为《由经典和EPR信道传送未知量子态》的论文，率先提出量子隐形传态的设想。论文提出的方案是：将甲地量子态所含的信息分为经典信息和量子信息两部分，分别由经典信道和量子信道（利用量子纠缠实现）送到乙地，接收者在获得这两种信息后，在乙地重新构造出甲地量子态的原貌。这种“隔空传态”的设想立刻引起了人们的兴趣，因为从某种意义上来说，“隔空传态”和“隔空传物”的效果是一样的，新的粒子和原来的粒子一模一样，那不就相当于把原来的粒子传送过去了吗？

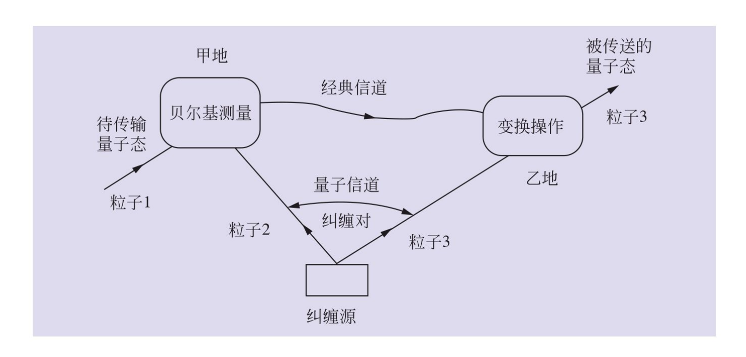



图28-1　量子隐形传态原理示意图

量子隐形传态的原理如图28-1所示。粒子1是甲地需要传态的原物粒子，处于某种未知的量子态。粒子2和粒子3是一对处于纠缠态的粒子，分别发送至甲地和乙地，由于粒子2和粒子3处于纠缠态，因此只要一方被测量，另一方会瞬时发生相应的变化。然后，在甲地对粒子1和粒子2进行一种叫做贝尔态分析的联合测量。在贝尔基测量过程中，粒子1与粒子2随机地以四种可能方式之一纠缠起来，导致3个粒

子之间实现了“纠缠转移”，粒子1原来量子态的大部分信息转移到了粒子3上。然后，甲把贝尔基测量结果通过经典信道告诉乙，乙便获得了剩余的信息，于是可以采取相应的操作，将粒子3转换成粒子1原来的量子态。这就是量子隐形传态的全过程。在此过程中，发送者对粒子1的量子态一无所知，贝尔基测量完成后，粒子1的量子态就被破坏了。需要注意的是，由于量子隐形传态需要借助经典信道才能实现，因此并不能实现超光速通信。

量子隐形传态方案提出以后，科学家们纷纷开始尝试实验验证。1997年，塞林格团队率先成功，他们将一个光子的未知偏振态利用量子隐形传态成功传输至另一个光子上，该实验直观地向人们展示了量子力学的神奇，引起了巨大轰动。随后，世界各国的科学家们如火如荼地开展了各种量子隐形传态实验。量子隐形传态又先后在冷原子、离子阱、超导、量子点和金刚石色心等诸多物理系统中得以实现。

量子隐形传态能够借助量子纠缠将未知的量子态传输到遥远地点，而不用传送物质本身，因而可以作为一种简单而又神奇的量子通信方式来传输量子比特。量子隐形传态是远距离量子通信和分布式量子计算的核心功能单元，在量子通信和量子计算网络中发挥着至关重要的作用。

作为塞林格团队的主要成员，潘建伟深受塞林格器重。塞林格曾经问过潘建伟一个问题，你的梦想是什么？潘建伟的回答是：在中国建设一个世界一流的量子物理实验室。回国后，他做到了。2006年，潘建伟团队首次实现两个光子的偏振态隐形传态；2015年，他的团队又成功地实现了多自由度的量子隐形传态，该成果被英国物理学会新闻网站“物理世界”评选为“国际物理学年度突破”。

量子隐形传态是量子纠缠的重要应用，但是，量子纠缠却有一个致命的缺点——量子纠缠十分脆弱，环境的退相干作用会不可避免地破坏其量子特性而使“纠缠”消失掉，即两个纠缠的量子客体最终会演化为不纠缠的状态。环境的退相干作用不仅包括经典噪声，诸如热运动、电磁场、吸收、散射等，还包括量子噪声，即真空量子涨落（真空能量波动导致真空中不断地有各种正反虚粒子对产生并迅速湮灭）。即使你能将经典噪声完全隔绝，量子噪声也无法消除，而且无

处不在。因此，如何采取措施克服退相干，拓展量子隐形传态的传输距离，是一个重要的研究课题。

光纤中的损耗和退相干效应比较显著，因此隐形传态的距离受到了极大的限制。2004年，塞林格团队利用多瑙河底的光纤信道，成功地使量子隐形传态的距离达到了600 m。2020年，美国加州理工学院的研究团队在光纤信道内实现了44 km的远距离量子隐形传态，保真度大于90%。

2004年，潘建伟团队开始探索在自由空间中实现更远距离的量子通信。自由空间简单来说就是没有物质的空间，如外太空。在自由空间，环境对光子的干扰极小，光子一旦穿透大气层进入外层空间，其损耗便接近于零，这使得光纤在自由空间比远距离传输方面更具优势。2012年，潘建伟团队在青海湖上空首次成功实现了百千米级的自由空间量子隐形传态。2017年，借助“墨子号”量子科学实验卫星，该团队成功实现长达1400 km的量子隐形传态，创造了传输距离的世界纪录。

上面介绍的单光子偏振态的量子隐形传态属于离散变量方式，量子隐形传态还有一种方式叫连续变量量子隐形传态。离散变量实验中所使用的是一个一个的单光子，而在连续变量实验中，以由大量光子组成的光学模为基本单元，其探测效率要比离散变量更高。1998年，美国加州理工学院首次实现了连续变量的量子隐形传态。2016年，我国山西大学光电研究所在国际上首次实现了长达6 km距离的基于光纤的连续变量量子隐形传态。

量子隐形传态最容易引起人们遐想的地方，莫过于它是否可以实现“隔空传物”甚至“隔空传人”。毕竟，人也是由微观粒子组成的，尽管数量大到近乎天文数字。其设想是，是否可以把一个人身上所有粒子的量子信息传递到另一地的粒子上进行人体重组？这个设想已经完全超出了现阶段物理学家们的能力，实现的可能性为零。但是，假如说在遥远的未来真的实现了“隔空传人”，按照量子隐形传态原理，必须把一个人在一地摧毁，然后才能在异地重建，那么，即使重建的人和被摧毁的人完全一样，他还是原来的他吗？

# 29 展望未来

量子力学是一场科学上的革命，它几乎颠覆了以牛顿力学为代表的经典物理的所有观念，让人类对世界的认识提高了一个层次。

同时，量子力学也给人类带来了技术上的革命。第一次量子革命催生的相关技术早已深入到我们日常生活的每个角落，在这些技术里，量子力学隐身幕后，深藏功与名，如激光、半导体、晶体管、核磁共振、高温超导、原子钟等。这些器件功能上遵从经典物理规律，但其运行基础却是基于量子力学原理，如果没有量子力学，人类就无法研究其物理原理，也就很难发明出这些技术。我们习以为常的各种芯片离不开晶体管，卫星导航离不开原子钟，可以说，正是第一次量子革命，才使人类进入了现代信息社会。

随着量子信息技术的开发，量子力学从幕后走到了台前，带来了第二次量子革命。量子通信、量子计算、量子密码、量子网络、量子模拟、量子传感、量子雷达、量子导航、量子关联成像、量子精密测量等技术，令人目不暇接，眼界大开。这些量子器件和技术在功能上直接遵从量子力学规律，可以完成经典技术所不能完成的任务。这些崭新的技术将会给人类社会再一次带来翻天覆地的变化。

在量子力学的世界里，量子态的叠加（相干性）、纠缠（非定域性）和测量（随机性）是其区别于经典力学的最主要的特性，也是各种量子信息器件的技术基础。同时，量子态的不可克隆性是这些技术的安全基础。反过来，环境的退相干效应则是这些技术需要面对的主要问题。

近年来，世界各国纷纷推出了量子信息技术的国家战略，力争把握第二次量子革命的历史机遇，争做量子信息时代的领头羊。可喜的是，我国在这一次科技浪潮中牢牢把握住了机遇，在量子通信领域处于国际领先地位，在量子计算领域与发达国家整体处于同一水平线。

现阶段，量子科技的国际竞争日益激烈，技术发展日新月异。以

量子计算机为例，2019年，美国谷歌公司发布53比特超导量子计算机“悬铃木”，宣称实现“量子霸权”；2021年，我国发布了62比特超导量子计算机“祖冲之号”和66比特的“祖冲之二号”，量子比特数目超过了“悬铃木”；而到了2021年年底，IBM公司宣称已经研制出了一台能运行127个量子比特的超导量子计算机“鹰”，再次打破纪录。你追我赶，争夺异常激烈。

而在技术路线上，创新也是层出不穷。本书前面提到的量子计算都是基于量子逻辑电路，与经典的图灵机具有类似的架构，可以称之为标准量子计算。近年来，一些科学家对如何实现量子计算提出了一些不同的架构，如拓扑量子计算、绝热量子计算（量子退火算法）、单向量子计算等，这些量子计算架构具有退相干时间长、抗干扰能力强等优点。在这些新的设想中，绝热的量子退火计算机发展最快。量子退火计算机的主要用途是求解某些最优化问题，它执行的是量子退火算法，这是一种利用量子波动产生的量子隧道效应来搜寻问题最优解的算法。加拿大的D-Wave公司在该领域处于世界领先地位，已经推出了商业化产品（图29-1）。2017年，D-Wave公司推出由2000个比特构成的超导量子退火计算机，它的处理器由排列于整齐格子中的金属铌超导线圈构成，每个线圈是一个量子比特，在接近绝对零度的温度下工作，对于最优化问题，该机胜过当前高度专业化的经典算法1000～10 000倍。2020年，该公司又发布了5000量子比特的退火计算机，再次刷新纪录。

过去100年来，第一次量子革命从根本上改变了人类的生活方式。我们有理由相信，在未来的100年，第二次量子革命还会创造更多的奇迹，让我们做好准备，一起迎接这激动人心的新时代吧！



图29-1　D-Wave公司推出的量子退火计算机

(https://baike.baidu.com/item/D-Wave%20 2X/18698843)

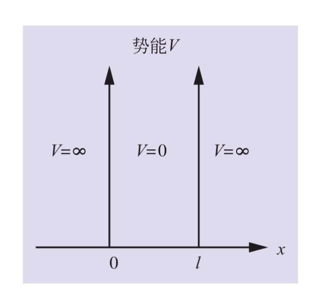
# 附录

## 附录A　一维无限深势阱中自由粒子的运动

薛定谔方程在量子力学中的作用，相当于牛顿方程在经典力学中的作用。处理量子力学问题，首先就是写出薛定谔方程，然后进行求解，可解得能量与波函数，进而可求其他可观测量，最后对解的结果进行分析与讨论。

薛定谔方程的求解在多数情况下是很困难的，只有少数几个例子是可以精确求解的。下面我们就来看一个可以精确求解的例子——一维无限深势阱中自由粒子的运动。通过对薛定谔方程的求解，我们可以认识到许多奇异的量子特性。

一维无限深势阱中自由粒子是指：一个质量为$m$的粒子，沿$x$轴在一维方向上运动，它受到如图A-1所示的势能的限制（图中纵坐标表示势能，由于这个图像一个井，所以被称为势阱），阱外势能无穷大、阱内势能为零。由于阱外势能无穷大，故该粒子在阱外永不出现；而阱内势能为零，故该粒子在阱内不受力而作自由运动。也就是说，该粒子被限制在$x$轴上0～$l$范围内自由运动。




图A-1　一维无限深势阱

该粒子的薛定谔方程为

$$-\frac{\hbar^2}{2m}\frac{\mathrm{d}^2\psi(x)}{\mathrm{d}x^2}=E\psi(x) \quad \text{(A-1)}$$

这是一个微分方程，其求解超出了本书的范围。读者可以参考相关量子力学教科书，本书直接给出求解结果。

粒子的能量：

$$E_n=\frac{n^2h^2}{8ml^2} \quad \text{(A-2)}$$

粒子的波函数：

$$\psi_n(x)=\sqrt{\frac{2}{l}}\sin\frac{n\pi x}{l}(0\leqslant x\leqslant l) \quad \text{(A-3)}$$

上面两个式子里的$n$是在求解过程中自然引入的参数，$n$只能取正整数（$n=1, 2, 3, 4,\ldots$），称之为量子数。（A-2）式中的$h$是普朗克常数。

## 1．能量

首先来对能量进行分析。由（A-2）式可以看出，由于$n$只能取正整数，所以粒子的能量只能取一些离散的数值，这就是量子力学的重要特性——能量量子化。这里量子化的得出是由薛定谔方程“自然地”得到的，而不像普朗克和玻尔那样是人为“强加”给粒子的。这样量子力学对能量量子化的解释就更为合理和顺畅，也使人们更容易判断什么情况下能量是量子化的，什么情况下可以近似看作是连续的。我们来看下面几个例子。

已知两个能级的能量差$\Delta E_n=E_{n+1}-E_n$，求下面三种情况下$\Delta E_n=$？

例1：$m=9.11\times10^{-31}$ kg的电子，在$l=10^{-10}$ m的一维势阱中；

例2：$m=9.11\times10^{-31}$ kg的电子，在$l=0.01$ m的一维势阱中；

例3：$m=10^{-3}$ kg的粒子，在$l=1$ m的一维势阱中。

代入公式计算，可以得到如下结果。

例1：$\Delta E_n=(2n+1)\times 38$ eV；

例2：$\Delta E_n=(2n+1)\times 2.35\times 10^{-15}$ eV；

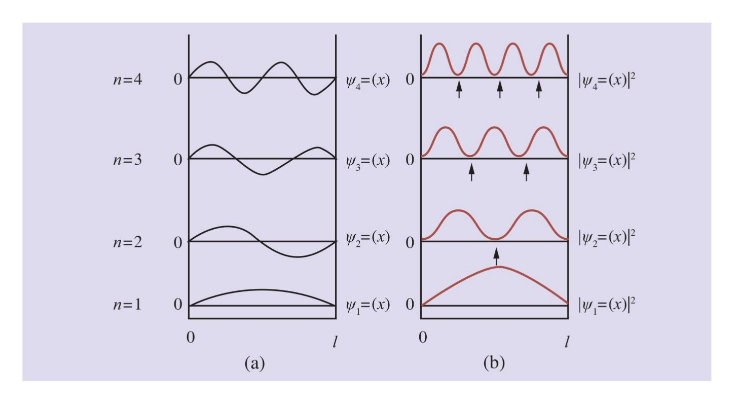
例3：$\Delta E_n=(2n+1)\times 3.43\times 10^{-46}$ eV。

对于例1，相对于电子这样的微观粒子来讲，能级间隔非常大，能量是量子化的。对于例2，能级间隔非常小，可以近似认为能量是连续的。$10^{-10}$ m是原子尺度，0.01 m是宏观尺度，也就是说，如果电子在原子尺度内运动，量子化特征非常明显；但是，如果它在宏观尺度内运动，量子化特征基本消失。正因为如此，原子中电子的运动由于量子特性而让人捉摸不定，但电视机显像管中的电子又能在荧光屏上呈现出我们想要的图像，而不是一团乱麻。

通过例3，可以看到如果是一个宏观粒子，由于$m$和$l$都很大，所以能级间隔小到没有意义，能量完全可以看成是连续的，已经完全失去了量子特性。

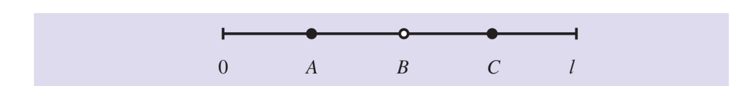
## 2．波函数

接下来对波函数进行分析。波函数的模的平方$|\psi(x)|^2$具有明确的物理意义：$|\psi(x)|^2$表示在坐标$x$点发现粒子的概率密度。图A-2给出了由（A-3）式绘制的$\psi(x)$和$|\psi(x)|^2$的图像。可以看出，波函数$\psi(x)$是一种正弦波图像，而且能量越高，其“振动”越剧烈。在对波函数进行分析时，最有意义的是$|\psi(x)|^2$，它给出了粒子的空间概率密度分布图像，对于一维的$x$轴，事实上就体现出粒子在这条轴上每一点出现的概率。从$|\psi(x)|^2$图像可看出，当此粒子处于基态时（$n=1$），粒子在$l/2$处出现的概率最大；当粒子处于第一激发态时（$n=2$），在$l/4$和$3l/4$两处出现的概率最大，但是在$l/2$处出现的概率为零，我们把概率为零这一点叫做节点。可以看到，$n$越大，节点数越多。




图A-2　一维无限深势阱中粒子的ψ(*x*)和|ψ(*x*)|²图像，|ψ(*x*)|²图中箭头所指的位置为节点

（a）波函数；（b）概率密度




图A-3　一维无限深势阱中粒子处于第一激发态的运动

从经典力学的角度来看，存在节点是不可想象的。为什么这么说呢？我们来分析一下第一激发态（$n=2$）粒子的运动。如图A-3所示，粒子在$x$轴上0~$l$范围内做一维运动，中心的$B$点是节点，粒子在$B$点左右两边都有出现的概率，但在$B$点出现的概率为零。那么问题来了，如果粒子在$A$点出现以后又在$C$点出现，那么它是怎么过去的？

按经典力学，从$A$点到$C$点，粒子只能沿着轴移动过去，但这样就必然会经过$B$点，那么$B$点的概率就不为零，它就不再是节点。节点的存在，意味着量子运动和经典运动是完全不同的，粒子可以从$A$点到$C$点，但是并不经过$B$点，我们没法想象粒子的运动轨迹，唯一合理的解释就是：它没有运动轨迹！

粒子没有固定的运动轨迹，只有概率分布的规律，这是量子力学中粒子运动的普遍规律，事实上，这也是量子力学中不确定关系（见第7章）的必然结果，如果有轨迹，动量和位置就同时确定了，就不满足不确定原理了。

## 3．零点能

再来审视一维无限深势阱中粒子的能量。对于（A-2）式，由于$n$是正整数，我们发现粒子的能量有一个最小值，即$n=1$时的能量$E_1$，且$E_1>0$。由于势能为0，则$E_1$为粒子的动能，可见粒子的动能恒大于0，这就是零点能效应。零点能效应表明粒子是无法静止的，这和经典力学完全不同，因为经典粒子是可以静止的，动能可以为零。事实上，零点能效应也是量子力学中不确定关系的必然结果，如果静止，动量和位置就同时确定为零，那就违反了不确定原理。

零点能效应使人们对于物体降温到绝对零度时会不会完全静止有了正确的认识。

我们知道，温度是反映物体分子热运动的一个物理量。物体内部的原子和分子都在运动。运动越剧烈，温度越高。显然，当一个物体降温的时候，它的分子运动速度越来越慢，当达到最慢速度的时候，温度就达到了最低值，也就是绝对零度，它等于 - 273.15℃。物体降温的时候，会由气体变成液体再变成固体，因为气体的分子热运动是最快的，固体是最慢的。如氧气，在降温的时候，它会先变成液体再变成固体，都是淡蓝色，非常漂亮。

那么，当物体降温到绝对零度时，它的内部粒子是不是就完全不动了呢？其实不是。根据薛定谔方程的计算，固体晶格振动的能量是量子化的（见第22章），固体在绝对零度的时候，内部晶格振动能量达到一个最低值，这就是零点能。这时候粒子振动非常微弱，但是不为零。

虽然几乎一切物质在绝对零度时都会变成固体，但有一种物质例外，那就是氦。氦的零点能比较大，即使降温到绝对零度，它也不会固化，仍然保持液态，所以人们把氦叫做永久液体。正是利用氦的这一特性，人们研制出了氦制冷机来获得极低的温度。大型氦低温制冷机是超导、核聚变、高能物理等前沿科技研究中不可或缺的基础设备。

氦还有一个特殊的性质，就是当它接近绝对零度的时候，会变成

超流体。超流体非常神奇，如果你把超流体放在杯子里，它会自动沿着杯壁往外爬，直到流完为止；超流体还能丝毫不受阻滞地流过管径极细的毛细管。研究表明，液氦从正常相变成超流相时，液体中的原子会突然失去随机运动的特性，而以整齐有序的方式运动。于是，液氦失去了所有的内摩擦力，它的热导率会突然增大100万倍，黏度会下降100万倍，从而使它具有了一系列不同于普通流体的奇特性质。

## 附录B　氢原子中电子的运动

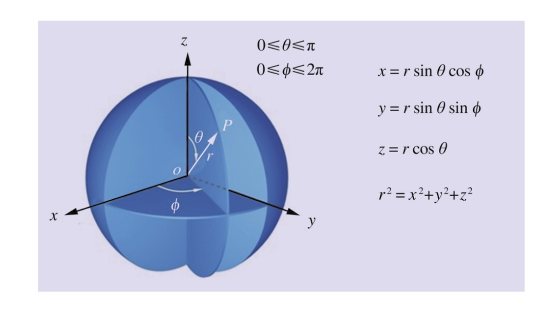
量子力学使人们对物质结构有了本质的理解。氢原子是最简单的原子，也是唯一一个能够精确求解其薛定谔方程的原子，正是从它身上，薛定谔揭开了原子结构的奥秘。

考虑到原子核质量远远大于电子质量，我们假设原子核不动，然后通过薛定谔方程来求解电子绕核运动的规律。氢原子的薛定谔方程如下：

$$\left[-\frac{\hbar^2}{2m_e}\left(\frac{\partial^2}{\partial x^2}+\frac{\partial^2}{\partial y^2}+\frac{\partial^2}{\partial z^2}\right)-\frac{e^2}{4\pi\varepsilon_0 r}\right]\psi(x,y,z)=E\psi(x,y,z)$$

式中，$m_e$为电子质量；$e$为电子电量；$\varepsilon_0$为真空介电常数；$r$为电子离核距离。

为了能够求解方程，需要把直角坐标（$x$，$y$，$z$）变换为球极坐标（$r$，$\theta$，$\varphi$）。以原子核为坐标原点，假设电子在直角坐标系的位置为点$P$（$x$，$y$，$z$），那么$P$点到原点$O$的距离就是$r$，$OP$连线与$z$轴的夹角就是$\theta$，连线在$xy$平面内的投影与$x$轴的夹角就是$\varphi$。两种坐标系的变换关系见图B-1。




图B-1　直角坐标系与球极坐标系的变换关系

氢原子的薛定谔方程求解过程相当复杂，本书仍然直接给出求解结果。在求解过程中，自然引入了3个量子数，分别是主量子数$n$、角量子数$l$和磁量子数$m$。

## 1．能量量子化

求解得到电子的能量为

$$E_n=-\frac{13.6}{n^2}\ \mathrm{eV}\ (n=1, 2, 3, \cdots) \qquad (\mathrm{B-1})$$

能量取负值是因为将电子离核无穷远时的势能定为0。可以看出，能量是量子化的，$n$越大，电子能级越高。

## 2．波函数

电子的波函数$\psi_{nlm}(r,\ \theta,\ \varphi)$表达式很复杂，不同的$n$、$l$、$m$对应不同的波函数，用不同的下标标记。

当$n=1$时，$E_1=-13.6$ eV，此时波函数有1个解（$\psi_{1s}$）；

当$n=2$时，$E_2=-3.40$ eV，此时波函数有4个解（$\psi_{2s}$、$\psi_{2px}$、$\psi_{2py}$、$\psi_{2pz}$）；

当$n=3$时，$E_3=-1.51$ eV，此时波函数有9个解……每一个能级$E_n$对应$n^2$个波函数。表B-1给出了几个低能级波函数的表达式。

表B-1　氢原子中的电子波函数$\psi_{nlm}$（$r, \theta, \varphi$）

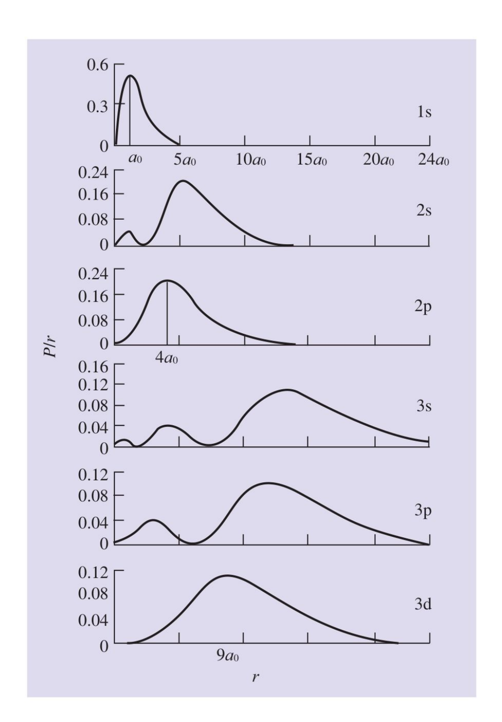
| 量子数取值 | | | 波函数 $\psi_{nlm}(r,\theta,\phi)$ | 波函数命名 |
|---|---|---|---|---|
| $n$ | $l$ | $m$ | | |
| 1 | 0 | 0 | $\psi_{1s}=\sqrt{\frac{1}{\pi}}\left(\frac{1}{a_0}\right)^{\frac{3}{2}}e^{\frac{r}{a_0}}$ | 光谱上将 $l=0, 1, 2, 3, \cdots$记为 s, p, d, f, ⋯ 故 $n=1$、$l=0$ 记为 1s |
| 2 | 0 | 0 | $\psi_{2s}=\sqrt{\frac{1}{32\pi}}\left(\frac{1}{a_0}\right)^{\frac{3}{2}}e^{-\frac{r}{2a_0}}\left(2-\frac{r}{a_0}\right)$ | $n=2$、$l=0$ 记为 2s |
| | 1 | 0 | $\psi_{2p_z}=\sqrt{\frac{1}{32\pi}}\left(\frac{1}{a_0}\right)^{\frac{5}{2}}e^{-\frac{r}{2a_0}}r\cos\theta$ | $n=2$、$l=1$ 为 2p，$r\cos\theta=z$，记为 $2p_z$ |
| | 1 | ±1 | $\psi_{2p_x}=\sqrt{\frac{1}{32\pi}}\left(\frac{1}{a_0}\right)^{\frac{5}{2}}e^{-\frac{r}{2a_0}}r\sin\theta\cos\phi$ | $n=2$、$l=1$ 为 2p，$r\sin\theta\cos\phi=x$，记为 $2p_x$ |
| | | | $\psi_{2p_y}=\sqrt{\frac{1}{32\pi}}\left(\frac{1}{a_0}\right)^{\frac{5}{2}}e^{-\frac{r}{2a_0}}r\sin\theta\sin\phi$ | $n=2$、$l=1$ 为 2p，$r\sin\theta\sin\phi=y$，记为 $2p_y$ |

注：$a_0=52.9$ pm，称为玻尔半径。

## 3．电子云

我们已经知道，波函数模的平方$|\psi|^2$代表在空间某点发现粒子的概率密度。所以我们将$|\psi_{nlm}(r,\ \theta,\ \varphi)|^2$作图，就能看出电子在原子核周围空间的概率密度分布。$|\psi|^2$函数图形就是“电子云”，如图5-2所示。

仔细观察1s轨道的电子云，会发现颜色最深的地方在原子核上，这是不是意味着电子在核上出现的概率最大呢？并不是！这是一个常见的误区，就是把概率密度和概率混淆，事实上，这是两个不同的概念。概率密度是单位体积内电子出现的概率。要想知道电子在某一点出现的概率，需要用该点的概率密度乘以该点的体积，这就要用微积分来处理——$|\psi|^2d\tau$表示在空间某一点附近微体积元$d\tau$内发现电子的概率，把$|\psi|^2d\tau$在某一范围内积分，就能算出此范围内电子出现的概率。据此，人们计算出了电子在距离原子核某一距离球壳内出现的概率，并将其作图，称为径向分布函数，如图B-2所示。从图中可以看出，对于1　s轨道，电子在距核$a_0$处出现的概率最大（$a_0=52.9$　pm，称为玻尔半径）；对于2p轨道，电子在距核$4a_0$处出现的概率最大；对于3d轨道，电子在距核$9a_0$处出现的概率最大。




图B-2　不同轨道的径向分布函数$P(r)$图（电子在距核$r$远、厚度为d$r$的球壳内出现的概率为$P(r)$d$r$）

对比图5-2和图B-2的1　s轨道电子，可以看到1　s电子在原子核上概率密度最大，但是这一点的球壳体积趋于零，所以电子在这一点出现的概率也接近零；随着离核距离$r$增大，概率密度在逐渐减小，但球壳体积在逐渐增大（球壳体积＝$4\pi r^2 \times$ d$r$）。经计算，两者乘积的极大值出现在离核52.9 pm处，这就是电子出现概率最大的地方。

## 4．节面

如果有一个粒子，它可以在篮球内部出现，也可以在篮球的外部出现，但是它在篮球球壳上出现的概率是0，那么这个篮球球壳就叫节面。节面最难理解的地方是，这个粒子从内到外或者从外到内，它是如何通过节面的？如果通过节面，节面的概率就不应该为0，那既然节面概率为0，它又是怎么进出的呢？千万不要以为这是无稽之谈，事实上，原子中的电子就是处于这样的运动状态。在电子云图

中，除1 s轨道外，其他轨道都有节面。

节面就是波函数$\psi=0$的面。因为$\psi=0$，所以电子在节面上出现的概率为零。电子云中有许多节面，例如，2 s轨道的节面是一个球面，3 s轨道的节面是两个球面（对应于图B-2中曲线与横轴的交点位置）……这也成为人们理解电子运动的难题之一，唯一的解释就是电子没有固定的运动轨迹，只有概率分布的规律。读者将节面的概念和一维无限深势阱中节点的概念进行比较，可以看到二者的物理内涵是一致的。

# 参考文献

[1]高鹏．从量子到宇宙——颠覆人类认知的科学之旅[M]．北京：清华大学出版社，2017．

[2]吴飙．简明量子力学[M]．北京：北京大学出版社，2020．

[3]井孝功，赵永芳．量子力学[M]．哈尔滨：哈尔滨工业大学出版社，2009．

[4]井孝功，郑仰东．高等量子力学[M]．哈尔滨：哈尔滨工业大学出版社，2012．

[5]曹天元．上帝掷骰子吗：量子物理史话[M]．沈阳：辽宁教育出版社，2008．

[6]张天蓉．群星闪耀：量子物理史话[M]．北京：清华大学出版社，2021．

[7]郭光灿，高山．爱因斯坦的幽灵：量子纠缠之谜[M]．2版．北京：北京理工大学出版社，2018．

[8]陈宇翱，潘建伟．量子飞跃：从量子基础到量子信息科技[M]．合肥：中国科学技术大学出版社，2019．

[9]袁岚峰．量子信息简话：给所有人的新科技革命读本[M]．合肥：中国科学技术大学出版社，2021．

[10]关洪．量子力学的基本概念[M]．北京：高等教育出版社，1990．

[11]费曼，莱顿，桑兹．费曼物理学讲义：第3卷[M]．潘笃武，李洪芳，译．上海：上海科学技术出版社，2020．

[12]费曼．QED：光和物质的奇妙理论[M]．张仲静，译．长沙：湖南科学技术出版社，2012．

[13]阿米尔·艾克塞尔．纠缠态：物理世界第一谜[M]．庄星来，译．上海：上海科学技术文献出版社，2016．

[14]魏凤文，高新红．仰望量子群星：20世纪量子力学发展史[M]．杭州：浙江教育出版社，2016．

[15]张三慧．大学物理学[M]．3版．北京：清华大学出版社，2017．

[16]周公度，段连运．结构化学基础[M]．5版．北京：北京大学出版社，2017．